Functional analysis of the serine-threonine protein kinase PknF and its substrate, the ABC transporter Rv1747, in *Mycobacterium tuberculosis*

Victoria Louise Spivey

A thesis submitted in partial fulfillment of the requirements of University College London for the degree of Doctor of Philosophy

Division of Mycobacterial Research
MRC National Institute for Medical Research
Mill Hill
London

December 2010

Declaration

I, Victoria Louise Spivey confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

V. L. Spivey

Acknowledgements

First and foremost I would like to thank my supervisor Roger Buxton for all his guidance and support over the last three years, without whom none of this work would have been possible. I would also like to thank my second supervisor Steve Smerdon for all his help and assistance with the protein and biochemical side of the project.

I would like to thank Debbie Hunt and Kathryn Lougheed for containment laboratory training. Thanks to Rachael Whalan and Juliet Curry for generation of the *ΔpknF* and *ΔRv1747* strains respectively. Thanks to Joanna Dillury for help with macrophage infections, Alan Williams for protein purification advice, Kristine Arnvig for seeing her Solexa traces and Lasse Stach for assistance in using the ITC machine. I would also like to acknowledge Liz Hirst for preparation of the electron microscopy samples and Stephen Howell for performing the mass spectrometry. I also thank Elaine Davis, Douglas Young and the rest of the Division of Mycobacterial Research for their valuable advice and discussions.

I acknowledge our collaborators Virginie Molle at the IBCP in France for help with pThr site identification; Angela Rodgers and Barry Walker at NIBSC for mouse infections; Del Besra and Apoorva Bhatt at the University of Birmingham for allowing me to spend time in their lab performing TLC experiments; Jake Bundy and Volker Beherends at Imperial College for performing NMR analysis and Ben Appelmelk and Jeroen Geurtsen at the VU University Medical Centre Amsterdam for kindly providing the LAM antibodies.

I owe special thanks to my family, friends and boyfriend all of whom have helped make this PhD an enjoyable experience.

Finally, I would like to thank the Medical Research Council for funding this studentship.

Abstract

Mycobacterium tuberculosis, the causative agent of tuberculosis exists in a number of different environmental states. It must therefore have gene regulatory systems which are specific for virulence. One major signalling method is through reversible phosphorylation of proteins, mediated by protein kinases and phosphatases. This study focuses on the function of one serine-threonine protein kinase, PknF, and its substrate, the ABC transporter Rv1747 which is necessary for growth in a virulent infection. This kinase is known to interact with both of the fork-head associated (FHA) domains of Rv1747 in a phosphorylation dependent manner. The aims of this study were to analyse the function of Rv1747 particularly in relation to its requirement for a virulent infection and to investigate how PknF is controlling Rv1747 function.

pknF was shown to be co-transcribed with *Rv1747* and the stimulus sensed by the kinase was investigated. Phenotypic analysis linked the function of Rv1747 to properties of the cell wall. Transcriptional microarray analysis of *pknF* and *Rv1747* mutants showed altered expression levels of genes involved in cell wall functions. Moreover, thin layer chromatography revealed changes in lipid profiles between wild type and mutant, but these differences could not be confirmed to be due to the mutation since they were not restored by complementation. Cell wall structure, however, appeared normal by transmission electron microscopy.

Experiments to determine how PknF regulates the function of Rv1747 demonstrated that phosphorylation occurs on two specific threonine residues; mutation analysis indicated that these are likely to be the only residues phosphorylated. The involvement of the FHA domains in this regulation was demonstrated by isothermal titration calorimetry, using peptides containing both phosphothreonine residues. Furthermore, FHA-1 domain mutation resulted in attenuation in macrophages highlighting its critical role in Rv1747 function. Infection experiments in macrophages and in mice have been performed using the threonine mutants to determine the *in vivo* consequences of phosphorylation and hence construct a model of how this regulation takes place. This study has revealed that PknF positively regulates the function of Rv1747 which is required for growth in both the lungs and spleens of mice.

Contents

Declaration **2**
Acknowledgements **3**
Abstract **4**
List of Figures **11**
List of Tables **16**
Abbreviations **17**

CHAPTER 1. INTRODUCTION **19**
1.1 Tuberculosis **19**
1.1.1 *Mycobacterium tuberculosis* complex 20
1.1.2 *M. tuberculosis* infection, treatment and vaccination 22
1.1.2.1 Infection and disease symptoms 22
1.1.2.1.1 *M. tuberculosis* infection cycle and immune response 24
1.1.2.2 Treatment 25
1.1.2.3 Vaccination 28
1.2 Protein Phosphorylation: A Universal Signal Transduction Pathway **30**
1.2.1 The protein kinase domain 31
1.2.2 *M. tuberculosis* signal transduction via protein phosphorylation 32
1.2.2.1 Histidine kinases and response regulators: the classical bacterial two-component system 35
1.2.2.2 The 'eukaryotic-like' serine-threonine protein kinases of *M. tuberculosis* 36
1.2.2.2.1 Structure, homology and activation of *M. tuberculosis* STPKs 36
1.2.2.2.2 *M. tuberculosis* kinases as novel drug targets 37
1.2.2.2.3 *M. tuberculosis* STPK function 39
1.3 The forkhead-associated domain **41**
1.3.1 FHA domain-containing proteins in *M. tuberculosis* 42
1.4 PknF and Rv1747 ..... **46**
1.4.1 ATP-binding cassette transporters 46
1.4.1.1 ABC transporters in *M. tuberculosis* 50
1.4.1.1.1 The ABC transporter Rv1747 51
1.4.2 Previous studies on the PknF-Rv1747 system 52
1.5 Project Hypotheses and Aims **55**
1.5.1 Hypothesis 56
1.5.2 Aims 56

CHAPTER 2. MATERIALS AND METHODS **57**
2.1 Bacterial Strains and Plasmids **57**
2.2 Recombinant DNA Techniques **61**
2.2.1 Design of oligonucleotides 61
2.2.2 Polymerase chain reaction (PCR) 61
2.2.3 Site directed mutagenesis (SDM) 62
2.2.4 Agarose gel electrophoresis 63
2.2.5 Plasmid DNA preparation 63

2.2.6 DNA fragment extraction and purification from agarose gels....63
2.2.7 Ethanol precipitation of nucleic acids....64
2.2.8 Restriction endonuclease digestion of DNA....64
2.2.9 Ligation of DNA....64
2.2.9.1 Ligation of DNA into pCR4Blunt-TOPO....64
2.2.10 Transformation of *E.coli*....64
2.2.11 Colony Screening and Glycerol Stocks....65
2.2.12 Determination of nucleic acid and protein concentration....65
2.2.13 DNA Sequencing....65
2.3 Transcriptional Analysis....66
2.3.1 Reverse Transcription PCR....66
2.3.2 5′ Rapid Amplification of cDNA Ends (5′ RACE)....66
2.3.3 Quantitative Real-Time PCR (qRT-PCR)....67
2.4 Mycobacteria Specific Techniques....68
2.4.1 Preparation and transformation of electrocompetent *M. tuberculosis*....68
2.4.1.1 Electroporation of mycobacteria....69
2.4.2 DNA isolation from *M. tuberculosis* on 7H11 agar plates and in liquid culture....69
1. Instagene Preparation of DNA....69
2. Genomic DNA isolation from *M. tuberculosis* for whole genome sequencing....69
2.4.3 RNA isolation from *M. tuberculosis* liquid cultures....70
2.4.4 Generation of *M. tuberculosis* cell free extract....71
2.4.4.1 Generation of *M. tuberculosis* protein extract for assessment of protein composition in culture supernatants....71
2.4.5 *M. tuberculosis* RNA-DNA microarrays....72
2.4.6 Lipid extraction and two dimensional thin layer chromatography of *M. tuberculosis* cell wall lipids....73
2.4.7 Transmission electron microscopy....75
2.4.8 ß-galactosidase reporter assay....75
2.4.9 *M. tuberculosis* growth inhibition (Alamar blue) assays....77
2.4.10 Metabolic tests for *M. tuberculosis*: Phenotype Microarrays (BiOLOG)....78
2.4.11 Analysis of small molecules and metabolites produced by *M. tuberculosis*.78
2.4.12 Murine bone marrow-derived macrophage infections....79
2.4.12.1 Differentiation of murine bone marrow derived macrophages (BMDMs)....79
2.4.12.2 Preparation of bacterial strains and infection of BMDMs....80
2.4.12.3 Greiss nitrite assay....81
2.4.13 Preparation of bacterial strains for infection in a mouse aerosol model....81
2.4.14 Enzyme-linked immunosorbent assays (ELISAs)....82
1. Cytokine ELISAs....82
2. *M. tuberculosis* whole cell ELISAs....83
2.5 Protein Expression and Analysis Techniques....83
2.5.1 Sodium dodoecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)....83
2.5.2 Expression and purification of *M. tuberculosis* proteins in *E. coli*....84
2.5.2.1 Protein expression....84

2.5.2.2 Protein Purification 85
2.5.2.2.1 Size exclusion chromatography 85
2.5.2.3 Mass spectrometry 86
2.5.3 Biochemical techniques 86
2.5.3.1 *In vitro* phosphorylation assays 86
2.5.3.2 Liquid chromatography / mass spectrometry / mass spectrometry (LC/MS/MS) 87
2.5.3.3 Isothermal titration calorimetry (ITC) 87
2.5.3.4 Size-Exclusion Chromatography Combined with Multiangle Laser Light Scattering (SEC-MALLS). 88

CHAPTER 3 RESULTS. CHARACTERISATION OF THE *pknF-Rv1747* OPERON, INVESTIGATION INTO THE *pknF* STIMULUS AND INITIAL ASSESSMENT OF THE *pknF* AND *Rv1747* MUTANT STRAINS. 91
3.1 Introduction 91
3.1 Experiments to determine co-transcription of *pknF* and *Rv1747* 92
3.2 Determination of the transcriptional start site of *pknF* 95
3.3 Investigation of the transcription of the *pknF-Rv1747* operon after stress 97
3.4 Assessment of the susceptibility of the *ΔpknF* and *ΔRv1747* mutants to a panel of drug and stress reagents. 102
3.5 Discussion 106
3.5.1 *pknF* and *Rv1747* are co-transcribed 106
3.5.2 The transcriptional start site of *pknF* 107
3.5.3 Investigation into the transcription of the *pknF-Rv1747* operon after stress 108
3.5.4 Assessment of the susceptibility of the *ΔpknF* and *ΔRv1747* mutants to a panel of drug and stress reagents. 109

CHAPTER 4 RESULTS. INVESTIGATION INTO THE NATURE OF THE Rv1747 SUBSTRATE. 111
4.1 Introduction 111
4.2 Transcriptional microarray studies of the *ΔpknF* and *ΔRv1747* mutant strains 114
4.2.1 The expression of many *M. tuberculosis* genes were affected by *Rv1747* deletion 115
4.2.2 The expression of only a few *M. tuberculosis* genes were affected by *pknF* deletion 118
4.2.3 The effect of the *pknF* and *Rv1747* deletions were confirmed for a selection of genes 123
4.3 Assessment of the lipid composition of the *M. tuberculosis* cell wall 129
4.3.1 Whole genome sequencing of wild type, *ΔRv1747* and *Rv1747* complementing strains of *M. tuberculosis* 134
4.4 Assessment of the lipoarabinomannan content of the mycobacterial cell wall 138
4.5 Transmission electron microscopy of the *M. tuberculosis* cell wall 141
4.6 Assessment of the protein composition of *M. tuberculosis* culture supernatants 141
4.7 Analysis of small molecules produced by *M. tuberculosis* 144

4.8 Metabolic tests for *M. tuberculosis*: Phenotype MicroArrays (BiOLOG) 146
4.9 Discussion....146
4.9.1 Transcriptional Microarrays and qRT-PCR....147
4.9.2 The search for the Rv1747 substrate....152

CHAPTER 5 RESULTS. HOW IS PknF CONTROLLING Rv1747 FUNCTION? PART 1.....161
ANALYSIS OF THE PHOSPHORYLATION DEPENDENT INTERACTIONS BETWEEN RECOMBINANT PknF AND Rv1747 *IN VITRO*.....161
5.1 Introduction....161
5.2 Expression and purification of PknF and Rv1747 from *E. coli*....162
5.3 Recombinant $PknF^{1-292}$ is an active kinase capable of autophosphorylation166
5.4 Autophosphorylation activity of $PknF^{1-292}$ is dependent on one specific lysine residue....166
5.5 $PknF^{1-292}$ phosphorylates $Rv1747^{1-559}$ *in vitro*....170
5.6 PknF phosphorylates Rv1747 *in vitro* on specific threonine residues....170
5.7 Construction of Rv1747 phosphorylation-null expression strains in *E. coli* 170
5.8 Expression and purification of $Rv1747^{1-559}$ threonine to alanine mutants in *E. coli*....173
5.8.1 Assessment of the phosphorylation state of the $Rv1747^{1-559}$ mutants....173
5.9 Assessment of the phosphorylation of $Rv1747^{1-559}$ by other *M. tuberculosis* serine-threonine protein kinases....176
5.10 Discussion....179
5.10.1 Protein purification....179
5.10.2 Autophosphorylation activity of recombinant PknF....181
5.10.2.1 The essential role of lysine 41 in PknF....183
5.10.3 Rv1747 and PknF and other kinase/substrate interactions....184
5.10.3.1 Identification and characterisation of Rv1747 phospho-threonine residues....185
5.10.3.1.1 Regulation of Rv1747 by phosphorylation....187
5.10.3.1.2 Mutation of Rv1747 residues T150 and T208....190
5.10.4 Rv1827 *in vitro* phosphorylation assays....192
5.10.5 *In vitro* phosphorylation assays....193
5.10.6 Summary....194

CHAPTER 6 RESULTS. HOW IS PknF CONTROLLING Rv1747 FUNCTION? PART 2.....195
FURTHER CHARACTERISATION OF THE *pknF-Rv1747* SIGNALLING SYSTEM AND ANALYSIS OF THE IMPORTANCE OF BOTH GENES FOR THE VIRULENCE OF *M. tuberculosis*.....195
6.1 Introduction....195
6.2 Assessment of the *Rv1747* null mutant in a murine macrophage infection model....196
6.3 Assessment of the growth of the *Rv1747* phosphorylation-mimic and phosphorylation-null mutants in macrophage and murine infection models 200
6.3.1 Construction of the panel of threonine mutations....200

6.3.2 Assessment of the growth of the *Rv1747* phosphorylation-mimic and phosphorylation-null double mutants in a macrophage infection model 201
6.3.3 Assessment of the growth of the *Rv1747* phospho-mimic and phospho-null double mutants in a murine aerosol infection model 203
6.3.4 Assessment of the growth of the *Rv1747* phospho-mimic and phospho-null single mutants in a murine macrophage infection model 203
6.4 Analysis of cytokine profiles.....207
6.5 Assessment of the growth of the *ΔpknF* and *pknF* complement strains in a macrophage model of infection model 211
6.6 Characterisation of the function of the Rv1747 FHA domains.....215
6.6.1 Assessment of the growth of the *Rv1747* complement S47A (ΔFHA-1) and *Rv1747* complement S248A (ΔFHA-2) strains in a murine macrophage infection model.....215
6.6.2 Expression and purification of FHA-1, FHA-2 and FHA-1 S47A domains of Rv1747 from *E. coli*.....217
6.6.3 Determination of the binding kinetics between the FHA-1 and FHA-2 domains with the Rv1747 pT150 and pT208 peptides 219
6.6.4 Determination of the binding kinetics between the FHA-1 and FHA-2 domains with the Rv1827 pT22 peptides.....220
6.6.5 Determination of the binding kinetics between the FHA-2 domain and the Rv0020c pT377 peptide 220
6.7 Does the FHA-2 domain dimerise upon phosphorylation at T208? 226
6.7.1 Determination of the phosphorylation of FHA-2 domain by PknF 226
6.7.1.1 *In vitro* phosphorylation of Rv1747, FHA-1 and FHA-2 domain by PknF 226
6.7.1.2 Use of electrospray mass spectrometry to determine the phosphorylation status of $FHA2^{202-310}$ 229
6.7.2 Use of SEC-MALLS to determine if the FHA-2 domain dimerises upon phosphorylation.....229
6.7.3 Expression and purification of GST-$PknF^{1-292}$ from *E. coli* 233
6.7.4 SEC-MALLS of FHA-2 phosphorylated by GST-$PknF^{1-292}$.....233
6.8 A hypothetical model of how PknF could be controlling Rv1747 function in *M. tuberculosis* 238
6.9 Discussion.....238
6.9.1 *ΔRv1747* is attenuated for growth in BMDMs.....238
6.9.1.1 Assessment of the growth of the threonine mutant strains in BMDMs and mice and implications for the regulation of the PknF-Rv1747 system.....241
6.9.1.1.1 BMDM infection……241
6.9.1.1.2 Murine infection……..242
6.9.1.1.3 BMDM infections with the single threonine mutants……244
6.9.1.1.4 Infection summary……..…245
6.9.1.2 Cytokine assays.....245
6.9.2 PknF 248
6.9.2.1 BMDM infections 248
6.9.3 Characterisation of the function of the Rv1747 FHA domains 250
6.9.3.1 BMDM infections 250
6.9.3.1.1 FHA domain function requires specific conserved residues..... 250

6.9.3.1.2 Conclusions drawn from Rv1747 FHA domain BMDM infections .. 250
6.9.3.2 Determination of the binding kinetics between FHA domains and pT peptides 252
6.9.3.2.1 Use of ITC to determine binding kinetics 252
6.9.3.2.2 FHA domain pT-containing epitope specificity..... 253
6.9.3.2.3 Principles of ITC and further explanation of results..... 254
6.9.3.3 Does the FHA-2 domain dimerise upon phosphorylation at T208? 256
6.9.4 Summary 258

CHAPTER 7. DISCUSSION AND FUTURE WORK 260
7.1 Discussion.....260
7.2 Future work and implications for novel *M. tuberculosis* treatment interventions 261

REFERENCES…… 264

APPENDIX……...278

List of Figures

Figure 1.1 The *M. tuberculosis* cell wall. 23

Figure 1.2 *M. tuberculosis* infection cycle. 26

Figure 1.3 New TB drugs under development..... 29

Figure 1.4 Diagram depicting the known interactions between the catalytic core of the protein kinase, ATP and a substrate protein. 33

Figure 1.5 A comparison of signalling systems in bacteria. 34

Figure 1.6 Structural analyses of *M. tuberculosis* serine/threonine protein kinases and protein phosphatase..... 38

Figure 1.7 *S. cerevisiae* RAD53 FHA-1 domain in complex with a phosphothreonine-containing peptide..... 43

Figure 1.8 Structural analyses of the *M. tuberculosis* FHA domain-containing proteins..... 44

Figure 1.9 Alignment of the *M. tuberculosis* FHA domains 45

Figure 1.10 The *pknF* and *Rv1747* genes plus the domain architecture of the two proteins..... 48

Figure 1.11 Typical structures of bacterial ABC transporters..... 49

Figure 1.12 Diagram of the genomic region of *M. tuberculosis* containing the *pknF* and *Rv1747* genes…….54

Figure 3.1 *pknF* and *Rv1747* are co-transcribed 94

Figure 3.2 The transcriptional start site of *pknF* 96

Figure 3.3 ß-galactosidase assay on the *pknF* promoter-*lacZ* strain and pEJ414 control strain in *M. tuberculosis*..... 98

Figure 3.4 ß-galactosidase assays on the *pknF* promoter-*lacZ* strain and pEJ414 control strain in *M. tuberculosis* after a panel of treatments. 101

Figure 3.5 Growth inhibition assays assessing the susceptibility of wild type H37Rv, *ΔpknF* and *ΔRv1747* strains to a range of drug and stress reagents. 104

Figure 4.1 qRT-PCR to confirm that transcription of *sigA* does not change between wild type, *ΔpknF*, *pknF* complement, *ΔRv1747* and *Rv1747* complement strains..... 126

Figure 4.2 qRT-PCR to confirm the effect of *Rv1747* deletion and complementation on the relative transcription levels of selected genes..... 127

Figure 4.3 qRT-PCR to confirm the effect of *pknF* deletion and complementation on the relative transcription levels of selected genes..... 128

Figure 4.4 Two dimensional thin-layer chromatography of wild type H37Rv, *ΔpknF* and *ΔRv1747* strains showing the apolar and polar lipids present in the cell pellets after separation in five different systems..... 131

Figure 4.5 Two dimensional thin-layer chromatography of wild type H37Rv, *ΔpknF*, *ΔRv1747*, *pknF* complement and *Rv1747* complementing strains showing the polar lipids present in the cell pellets after separation in system D..... 132

Figure 4.6 Two dimensional thin-layer chromatography of wild type H37Rv, *ΔpknF*, *ΔRv1747*, *pknF* complement and *Rv1747* complementing strains showing the polar lipids present in the culture supernatant of the *M. tuberculosis* strains after separation in system D..... 133

Figure 4.7 Summary of whole genome sequencing results..... 136

Figure 4.8 *M. tuberculosis* whole cell ELISAs comparing the levels of ManLAM in H37Rv wild type, *ΔRv1747* and *Rv1747* complement strains..... 139

Figure 4.9 Schematic structure of mannose-capped LAM (ManLAM)..... 140

Figure 4.10 Transmission electron micrographs of *M. tuberculosis* comparing cell wall structure in wild type H37Rv, *ΔpknF* and *pknF* complement strains..... 142

Figure 4.11 Transmission electron micrographs of *M. tuberculosis* comparing cell wall structure in wild type H37Rv, *ΔRv1747* and *Rv1747* complement strains..... 143

Figure 4.12 SDS-PAGE of proteins present in the supernatant of wild type *M. tuberculosis*, *ΔpknF*, *pknF* complement, *ΔRv1747*, and *Rv1747* complement strains. 145

Figure 5.1 Protein domains cloned for expression and purification from *E. coli*. 164

Figure 5.2 Expression and purification of recombinant $PknF^{1-292}$ and $Rv1747^{1-559}$ from *E.coli*..... 167

Figure 5.3. Recombinant $PknF^{1-292}$ is an active kinase capable of autophosphorylation. 168

Figure 5.4 A mutation of the critical lysine residue abolishes the autophosphorylation activity of PknF$^{1-292}$ 169

Figure 5.5 *In vitro* phosphorylation of Rv1747 by PknF. 171

Figure 5.6 Identification of phosphorylated threonine residues on Rv1747$^{1-559}$. 172

Figure 5.7 Expression and purification of wild type Rv1747$^{1-559}$, Rv1747$^{1-559}$ T150A, Rv1747$^{1-559}$ T208A and Rv1747$^{1-559}$ T150A/T208A. 174

Figure 5.8 *In vitro* phosphorylation of Rv1747$^{1-559}$ by PknF$^{1-292}$ is reduced in the threonine to alanine mutants. 175

Figure 5.9 Assessment of the phosphorylation of Rv1747$^{1-559}$ by other *M. tuberculosis* serine-threonine protein kinases. 177

Figure 5.10 Phosphorylation of Rv1827 by multiple kinases. 178

Figure 5.11 Phosphorylation site motif analysis for PknF. 188

Figure 5.12 Proposed mechanism for the interactions between PknF and Rv1747.... 189

Figure 6.1 Greiss nitrite assay after infection of BMDMs with wild type H37Rv, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. ..... …….198

Figure 6.2 Bone marrow derived macrophage infection with wild type H37Rv, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. 199

Figure 6.3 Bone marrow derived macrophage infection with wild type H37Rv, *ΔRv1747*, *Rv1747* complement, *Rv1747* complement T150A/T208A and *Rv1747* complement T150D/T208D *M. tuberculosis* strains. 202

Figure 6.4 Murine aerosol infection experiment with wild type H37Rv, *ΔRv1747*, *Rv1747* complement, *Rv1747* complement T150A/T208A and *Rv1747* complement T150D/T208D M. tuberculosis strains..... 205

Figure 6.5 Bone marrow derived macrophage infection with wild type H37Rv, *ΔRv1747*, *Rv1747* complement, *Rv1747* complement T150A, *Rv1747* complement T150D, *Rv1747* complement T208A and *Rv1747* complement T208D *M. tuberculosis* strains. 206

Figure 6.6 IL12-p40 production by BMDMs infected with wild type, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. 208

Figure 6.7 TNFα production by BMDMs infected with wild type, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. 209

Figure 6.8 IL6 production by BMDMs infected with wild type, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. 210

Figure 6.9 Bone marrow derived macrophage infection with wild type H37Rv, *ΔpknF* and *pknF* complement *M. tuberculosis* strains. 213

Figure 6.10 Bone marrow derived macrophage infection with wild type H37Rv, *ΔpknF* and *pknF* complement *M. tuberculosis strains* normalised to the six hour uptake data. 214

Figure 6.11 Bone marrow derived macrophage infection with wild type H37Rv, *ΔRv1747*, *Rv1747* complement S47A (ΔFHA-1) and *Rv1747* complement S248A (ΔFHA-2) *M. tuberculosis* strains..... 216

Figure 6.12 Expression and purification of FHA-1, FHA-2 and FHA-1 S47A domains of Rv1747 from *E. coli*. 218

Figure 6.13 Analysis of binding of Rv1747 FHA-1 domain to the Rv1747 pT150 and pT208 phosphopeptides. 221

Figure 6.14 Analysis of binding of Rv1747 FHA-2 domain to the Rv1747 pT150 and pT208 phosphopeptides. 222

Figure 6.15 Analysis of the binding of the Rv1747 FHA-1 S47A domain to the Rv1747 pT150 and pT208 phosphopeptides. 223

Figure 6.16 Analysis of binding of the Rv1747 FHA-1 and FHA-2 domains to the Rv1827 pT22 phosphopeptide. 224

Figure 6.17 Analysis of binding of the Rv1747 FHA-2 domain to the Rv0020c pT377 phosphopeptide..... 225

Figure 6.18 Schematic representation of how FHA-$2^{202-310}$ could potentially engage in an intermolecular interaction after phosphorylation at residue T208 227

Figure 6.19 *In vitro* phosphorylation of Rv1747, FHA-1 and FHA-2 by PknF. 228

Figure 6.20 Electrospray mass spectrometry of FHA-$2^{202-310}$ phosphorylated for four hours with $PknF^{1-292}$. 230

Figure 6.21 SEC-MALLS of FHA-2 domain with and without phosphorylation by PknF 232

Figure 6.22 Expression and purification of GST-$PknF^{1-292}$ from *E. coli*. 234

Figure 6.23 SEC-MALLS of the FHA-2 domain without phosphorylation by GST-PknF. 235

Figure 6.24 SEC-MALLS of the FHA-2 domain after phosphorylation by GST-PknF. 236

Figure 6.25 SEC-MALLS of the FHA-2 domain with and without phosphorylation by GST-PknF. 237

Figure 6.26. Schematic representation of a potential mechanism by which PknF controls Rv1747 function in *M. tuberculosis*. 239

List of Tables

Table 2.1. *E. coli* Strains used in this study 58

Table 2.2. Plasmids used and constructed in this study 59

Table 2.3 Antibiotics and supplements added to growth media. 60

Table 2.4. Solvents used to separate the different lipid components of the *M. tuberculosis* cell wall. 74

Table 2.5 The panel of drug and stress reagents used in this study to assess the promoter activity of *pknF* under different conditions. 77

Table 3.1 Summary of the minimum inhibitory concentrations derived from growth inhibition assays of wild type H37Rv, *ΔpknF* and *ΔRv1747* strains after treatment with a panel of drug and stress reagents. 104

Table 4.1 Microarray data for the genes whose differential expression levels were not restored upon complementation with the *Rv1747* gene..... 116

Table 4.2 Microarray data for the genes whose differential expression levels were not restored upon complementation with the *pknF* gene. 117

Table 4.3 Microarray data for the ten *M. tuberculosis* genes most highly up-regulated upon *Rv1747* deletion 119

Table 4.4 Microarray data for the ten *M. tuberculosis* genes most highly down-regulated upon Rv1747 deletion. 120

Table 4.5 Microarray data for the ten *M. tuberculosis* genes most highly up-regulated upon *pknF* deletion 122

Table 4.6 Microarray data for the ten *M. tuberculosis* genes most highly down-regulated upon *pknF* deletion 123

Table 4.7 Summary of SNPs identified from whole genome sequencing. 137

Abbreviations

aa	Amino Acid
ABC	ATP Binding Cassette
amp^R	Ampicillin Resistant
BCA	Bicinchoninic Acid
BCG	Bacille Calmette Guérin
BMDM	Bone Marrow Derived Macrophage
BμG@S	Bacterial Microarray Group at St. George's
bp	Base Pair
BSA	Bovine serum albumin
cDNA	Complementary DNA
CFE	Cell-Free Extract
CFU	Colony Forming Units
chlor^R	Chloramphenicol Resistant
cpm	Counts Per Minute
Da	Daltons
DE3	λDE3 Prophage
dCTP	2'-Deoxycytosine 5'Triphosphate
DEPC	diethylpyrocarbonate
dH_2O	Sterile Deionised Water
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-Deoxynucleoside 5'Triphosphate
DTT	Dithiothreitol
ELISA	Enzyme-linked Immunosorbent Assay
EDTA	Ethylenediaminetetraacetic Acid
FHA	Forkhead Associated
GSNO	S-Nitrosoglutathione
GSP	Gene Specific Primer
GST	Glutathione S-Transferase
HBSS	Hank's balanced salt solution
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulphonic Acid
HIV	Human Immunodeficiency Virus
hyg^R	Hygromycin Resistant
IFN-γ	Interferon-gamma
IPTG	Isopropyl-beta-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
Kan^R	Kanamycin Resistant
kb	Kilo Base
kDa	Kilo Dalton
kV	Kilo volts
l	litre
LAM	Lipoarabinomannan
LB	Luria-Bertani
L-cell	L-Cell Conditioned Media
LC/MS/MS	Liquid Chromatography/Mass Spectrometry/Mass Spectrometry

M	Molar
MALDI-TOF MS	Matrix-Associated Laser Desorption-Ionisation Mass Spectrometry
MOI	Multiplicity of Infection
mRNA	Messenger RNA
NEB	New England Biolabs
NO	Nitric Oxide
OD	Optical density
ONPG	Ortho-Nitrophenol-ß-D-Galactopyranoside
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
qRT-PCR	Quantitative Reverse Transcription PCR
rcf	Relative Centrifugal Force
RNA	Ribonucleic Acid
RNase	Ribonuclease
rpm	Revolutions Per Minute
RPMI	Roswell Park Memorial Institute Medium
RT	Reverse Transcriptase
RT-PCR	Reverse Transcription PCR
SEC-MALLS	Size Exclusion Chromatography-Multi Angle Laser Light Scattering
SDM	Site Directed Mutagenesis
SDS	Sodium Dodecyl Sulphafe
SSC	Saline-Sodium Citrate
STPK	Serine/Threonine Protein Kinase
TB	Tuberculosis
TLC	Thin Layer Chromatography
TBE	Tris-Borate EDTA
TdT	Terminal Deoxynucleotidyl Transferase
TE	Tris EDTA
TEMED	Tetramethylethlyethylenediamine
TOPO	DNA Topoisomerase I from *Vaccinia* Virus
TraSH	Transposon Site Hybridisation
U	Unit(s)
UV	Ultraviolet
V	Volts
X-gal	Bromo-chloro-indolyl-galactopyranoside

CHAPTER 1. INTRODUCTION

1.1 Tuberculosis

Tuberculosis (TB) remains one of the earliest recorded human diseases; clay tablets dating back to the seventh century B.C. describe people coughing blood and those from the fifth century B.C portray patients wasting away with chest pain and coughing (Smith, 2003). The causative agent of TB has been known to be the actinobacterium *Mycobacterium tuberculosis* since the description of the classic infection experiment by Robert Koch to the Berlin Physiological Society in 1882 (Brown, 1932; Smith, 2003). Subsequently, a live attenuated vaccine was developed at the Institute of Pasteur in Paris in the 1920's which has been given to more than three billion individuals worldwide (Andersen & Doherty, 2005; Behr & Small, 1999). Since the Second World War an effective drug therapy comprising multiple antibiotics has been in use and from the 1990's the World Health Organisation (WHO) has endorsed a global strategy called the DOTs (Directly Observed Treatment, Short-course) programme with the aim to properly monitor the treatment and outcome of each TB patient; consequently the programme has been identified as one of the most cost effective health strategies (http://www.tbalert.org/worldwide/DOTS).

It is therefore all the more remarkable that despite these preventative and therapeutic measures, TB has continued to be the world's most rampant infective agent, at the latest estimate killing 1.8 million people a year; furthermore, an estimated two billion people are latently infected with TB bacilli (WHO fact sheet 2009 (http://www.stoptb.org/)). Moreover there is apparently no animal host reservoir for *M. tuberculosis*. In 1993 the WHO declared TB to be a global health emergency to help raise awareness of the universal scale and magnitude of the problem with the aim to reduce the worldwide burden of the epidemic by 2015 (WHO fact sheet 2009). This disease burden now largely falls on developing countries; TB was effectively eradicated from Western Europe and North American countries in the 20th century, and re-occurrence in its incidence in the UK over the last 20 years are nearly all the result of immigration from

developing countries (Antoine *et al.*, 2006). Globally, this disease burden has escalated due to its deadly synergy with human immunodeficiency virus (HIV). TB is a leading cause of death among those people living with HIV; of the 1.8 million TB deaths in 2008, 0.5 million people were HIV positive (WHO fact sheet 2009). This HIV burden coupled with the emergence of multi-drug resistant (MDR) *M. tuberculosis* strains makes the search for novel vaccine candidates and drugs ever more important. In 2008 the cases of MDR-TB were the highest rates ever recorded (WHO fact sheet 2009).

Over the last 20 years TB has been a re-emerging problem in England and Wales; since 1988 the incidences of TB cases have increased year on year (Antoine *et al.*, 2006). In 2003, 45 % of all TB cases in England and Wales occurred in London (out of a total of 6780 cases) (Antoine *et al.*, 2006). The largest proportion of infections occurred in those people born outside the UK but infections were also prevalent in less affluent sections of the community including the homeless, drug users and prisoners (Antoine *et al.*, 2006). This highlights that TB is once again no longer only a problem of the developing world and it must be addressed to curb both the human suffering and enormous economic cost of this disease.

1.1.1 *Mycobacterium tuberculosis* complex

The *Mycobacterium* genus consists of rod shaped, high GC content, acid-fast Gram-positive bacteria. *Mycobacteria* are categorised as acid-fast due to the property of retention of a pink carbol-fuchsin stain when treated with acid-alcohol (Russell, 2001). This acid-fast property is due to the extremely hydrophobic cell wall of mycobacteria which contains molecules known as mycolic acids. The prevalence of these long chain fatty acids in the cell wall results in the acid-fast property as the fuchsin dye reacts with the carboxylic acid group of the mycolic acids (Madigan *et al.*, 2003).

The actinomycetes are a diverse group of Gram-positive bacteria which include *Mycobacterium*, *Corynebacterium*, *Nocardia* and *Streptomyces*. The genus *Mycobacterium* is thought to have originated in soil, like other actinomycetes, and then some species are believed to have evolved to colonise mammals (Smith, 2003). Mycobacteria can be divided into two groups according to *in vitro* growth rate;

organisms including *Mycobacterium smegmatis* are so called 'fast growers' as they take 3-4 days to form colonies whereas 'slow growers', which includes *M. tuberculosis*, take 3-6 weeks (Jacobs *et al.*, 1991).

The *M. tuberculosis* complex is comprised of seven members; *M. tuberculosis* is the primary cause of the disease in humans, *Mycobacterium bovis* can infect many animal species and up until the practice of pasteurisation, infected milk was the common route of infection into humans from this species (Cole, 2002). *Mycobacterium africanum* causes tuberculosis disease in humans in sub-Saharan Africa and *Mycobacterium cannetti* is a rare occurrence but also causes human TB. Apart from *M. bovis* other members of the complex which cause disease in animals include *Mycobacterium microti*, which infects seals and voles (Cole, 2002) and the newly identified *M. mungi*, which is the causative agent of TB among banded mongooses in Botswana (Alexander *et al.*, 2010).

It was commonly believed, up until a few years ago, that *M. bovis* was the evolutionary predecessor of *M. tuberculosis* (Smith, 2003). However, this hypothesis has now been considered doubtful as through analysis of the sequence insertions and deletions in the genomes of the *M. tuberculosis* complex it has been demonstrated that *M. bovis* and *M. tuberculosis* probably evolved independently from a common ancestor of *M. cannetti* (Brosch *et al.*, 2002).

M. tuberculosis H37Rv, a virulent and well-defined strain, was originally sequenced in 1998 (Cole *et al.*, 1998) and the genome was re-annotated four years later (Camus *et al.*, 2002). This strain was originally isolated from a 19 year old male TB patient in 1905 (Steenken & Gardner, 1946). The H37Rv genome is 4.4 megabases in size and contains approximately 4000 predicted protein encoding genes, about 40 percent of which encode proteins that have no known function (Sassetti *et al.*, 2003). Remarkably, approximately 18% of the annotated genes have predicted roles in cell wall and cell processes reflecting the incredible complexity of this structure in *M. tuberculosis* (Camus *et al.*, 2002).

M. tuberculosis has an unusual cell envelope, quite unlike that of typical gram-positive or gram-negative bacteria. Particularly characteristic is that it has a highly complex cell

wall rich in lipids, which make up to 60 percent of the dry weight of the bacteria (Kremer & Besra, 2005). The cell envelope consists of three main structures: the cytoplasmic membrane, the cell wall and the outer lipid layer (figure 1.1). The outer lipid layer functions as a second permeability barrier composed of mycolic acids and other lipid components (figure 1.1); in *M. tuberculosis* mycolic acids are critical for pathogenicity (Bhatt *et al.*, 2007). The extreme hydrophobicity of the cell wall accounts for its acid-fastness when stained with carbol-fuchsin, low permeability and resistance to many drugs and chemotherapeutics.

1.1.2 *M. tuberculosis* infection, treatment and vaccination

1.1.2.1 Infection and disease symptoms

Infection with *M. tuberculosis*, an obligate aerobe, occurs primarily via the respiratory route and is spread in the air through fine droplets containing the bacilli (Russell, 2001). In more then 90 percent of cases the infection is contained within organised collections of immune cells (macrophages and lymphocytes) which aggregate forming granulomas to halt disease progression (Russell, 2001). The ability to enter a latent-like state is one of the defining properties which makes *M. tuberculosis* a successful pathogen. Individuals with a latent infection have a 10 percent life time risk of re-activation whereas in individuals with a compromised immune system, such as those co-infected with HIV, the elderly, or those suffering from malnutrition, this risk increases to 10 percent per year (Russell, 2001). Progression to active pulmonary disease occurs when the granuloma breaks down and releases viable bacilli back into the airways of the lungs leading to aerosol spread of the infectious bacilli (Russell, 2001; Russell, 2007). Symptoms of a productive infection include a persistent cough, weight loss, fever, chest pain and night sweats (Smith, 2003). Extra-pulmonary tuberculosis can also develop through dissemination of the disease and can infect the lymphatic system, the central nervous system, the abdomen and bones; miliary TB is a severe form of the disease resulting from widespread dissemination of the disease to many organs of the body including the brain and is typified by multiple small lesions (Madigan *et al.*, 2003; Smith, 2003).

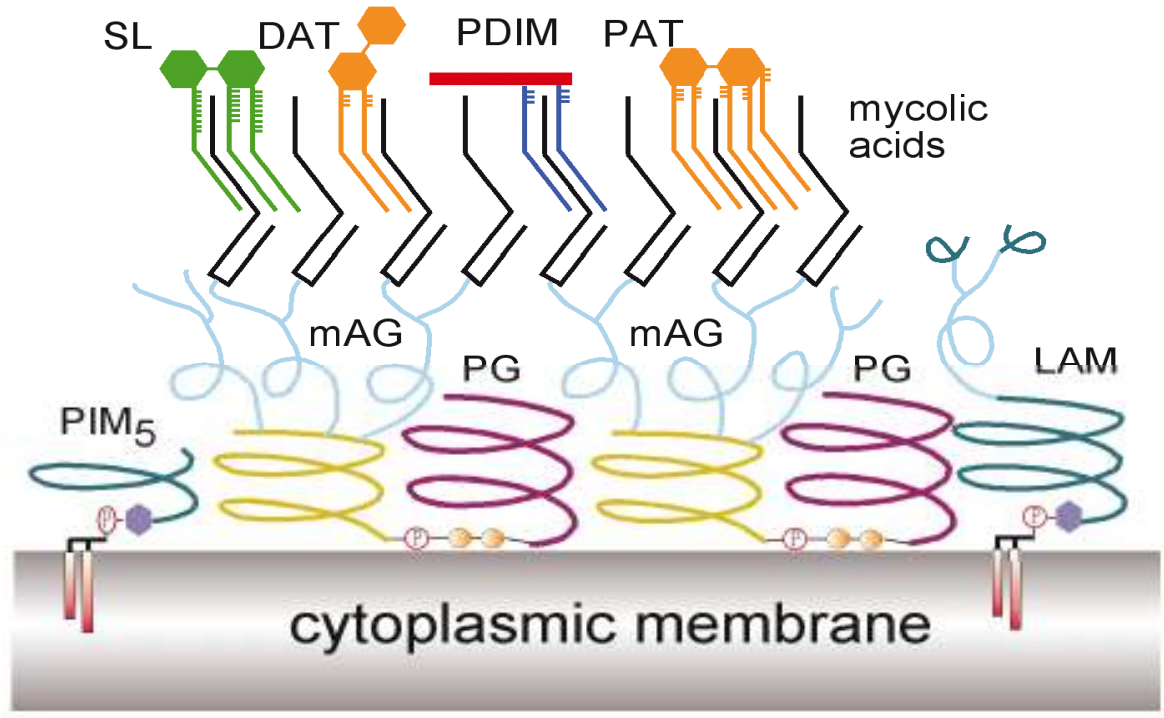


Figure 1.1 The *M. tuberculosis* cell wall. The complete structure is referred to as the mycolyl-arabinogalactan-peptidoglycan complex (Brennan & Nikaido, 1995). LAM = lipoarabinomannan, mAG = mycolyl arabinogalactan, PDIM = phthiocerol dimycoserosates, PG = peptidoglycan, SL = sulpholipid, DAT = diacyl trehalose, PAT = penta-acyl trehalose, PIM = phosphatidyl-myo-inositol mannoside (Kremer & Besra, 2005).

1.1.2.1.1 *M. tuberculosis* infection cycle and immune response

During the course of an *M. tuberculosis* infection the tubercle bacilli encounter changing environmental conditions. Upon inhalation, the bacilli enter the respiratory tract and are engulfed by alveolar macrophages lining the lungs (figure 1.2) (Russell *et al.*, 2010). Inside the macrophage, the *M. tuberculosis* bacilli are phagocytosed within vacuoles known as phagosomes. Usually, as phagosomes mature they fuse with lysosomes forming phagolysosomes; the acid hydrolase enzymes and lysosomal proteases within the lysosomes result in the digestion of the invading pathogen. However, *M. tuberculosis* has the ability to evade the host's defence mechanisms by inhibiting phagosome-lysosome fusion allowing survival and replication within the macrophages; it was demonstrated more than 40 years ago that phagosomes infected with *M. tuberculosis* failed to fuse with lysosomes (Armstrong & Hart, 1971).

Initial infection of alveolar macrophages leads to inflammation and as a result mononuclear cells (including lymphocytes, macrophages and monocytes) and neutrophils are recruited to the area from the blood vessels (figure 1.2) (Russell *et al.*, 2010); in addition, mycobacterial fragments can also be taken up by dendritic cells which migrate to the lymph nodes to activate T cells which in turn recruit and activate more macrophages (Stewart *et al.*, 2003). The collection of recruited cells forms the starting material for the granuloma which functions to wall off and contain the infected cells (Russell, 2007). As the granuloma matures it consists of infected macrophages surrounded by other macrophages which differentiate into several different cell types (figure 1.2). Lymphocytes are also recruited to granulomas and in later stages a fibrous wall of collagen is formed around the infected cells to produce a mature granuloma (Russell, 2007; Russell *et al.*, 2010). At this late stage the inside of the granuloma becomes hypoxic (Russell *et al.*, 2010). The *M. tuberculosis* infection can persist asymptomatically for an unknown length of time at this stage (Stewart *et al.*, 2003). Upon immune suppression (or anything which leads to the loss of CD4+ T cells) the granuloma breaks down releasing the caseated material facilitating a productive *M. tuberculosis* infection and the cycle can begin once more (Russell, 2007; Russell *et al.*, 2010).

The immune system plays a large role in protection against *M. tuberculosis* disease; a T_H1 cell mediated response is required for the defence against infection (Stewart *et al.*, 2003). *M. tuberculosis* can be eliminated when the macrophage is activated by a T_H1 cell (a subset of CD4+ T cells); T_H1 cells are first activated following antigen presentation by MHC class two molecules, T_H1 cells then secrete IFNγ which activates macrophages (Stewart *et al.*, 2003). Upon activation, macrophages increase their production of TNF receptors and are stimulated to secrete TNFα. A combination of TNFα secretion and the secretion of IFNγ by T_H1 cells induces the production of nitric oxide, superoxide ions, antimicrobial peptides and proteases which increases the antimicrobial action of the macrophage and helps to destroy the invading pathogen (Janeway *et al.*, 2005; Stewart *et al.*, 2003). IL-12 is produced by macrophages to induce CD4+ T-cell differentiation into T_H1-like cells; IL-12 is secreted by macrophages when mycobacteria ligate to Toll-like receptor 2 (TLR-2) on the surface of the macrophage, this ligation also induces the macrophage to produce intracellular nitric oxide (Janeway *et al.*, 2005). Activated macrophages are able to more effectively form phagolysosomes to destroy the invading pathogen (Janeway *et al.*, 2005; Stewart *et al.*, 2003).

Therefore, to allow the bacilli to persist inside macrophages requires complex regulation of gene expression; sensing the external environment allows *M. tuberculosis* to switch on and off appropriate sets of genes which facilitate survival. The identification of these virulence determinants may aid in the discovery of new drug targets and vaccine candidates which are needed to stem the burden of this disease.

1.1.2.2 Treatment

The discovery of streptomycin in the 1940's began the era of the antibiotic treatment of TB (Smith, 2003) and between 1952 and 1966 four new drugs were introduced: isoniazid (inhibitor of mycolic acid synthesis), pyrazinamide (unknown action), ethambutol (inhibitor of cell wall synthesis) and rifampicin (inhibitor of RNA synthesis) (Webb & Davies, 1999). Currently, treatment for active TB which is not drug resistant typically takes six months; initial treatment of a patient involves a four-drug combination of isoniazid, rifampicin, ethambutol and pyrazinamide for two months

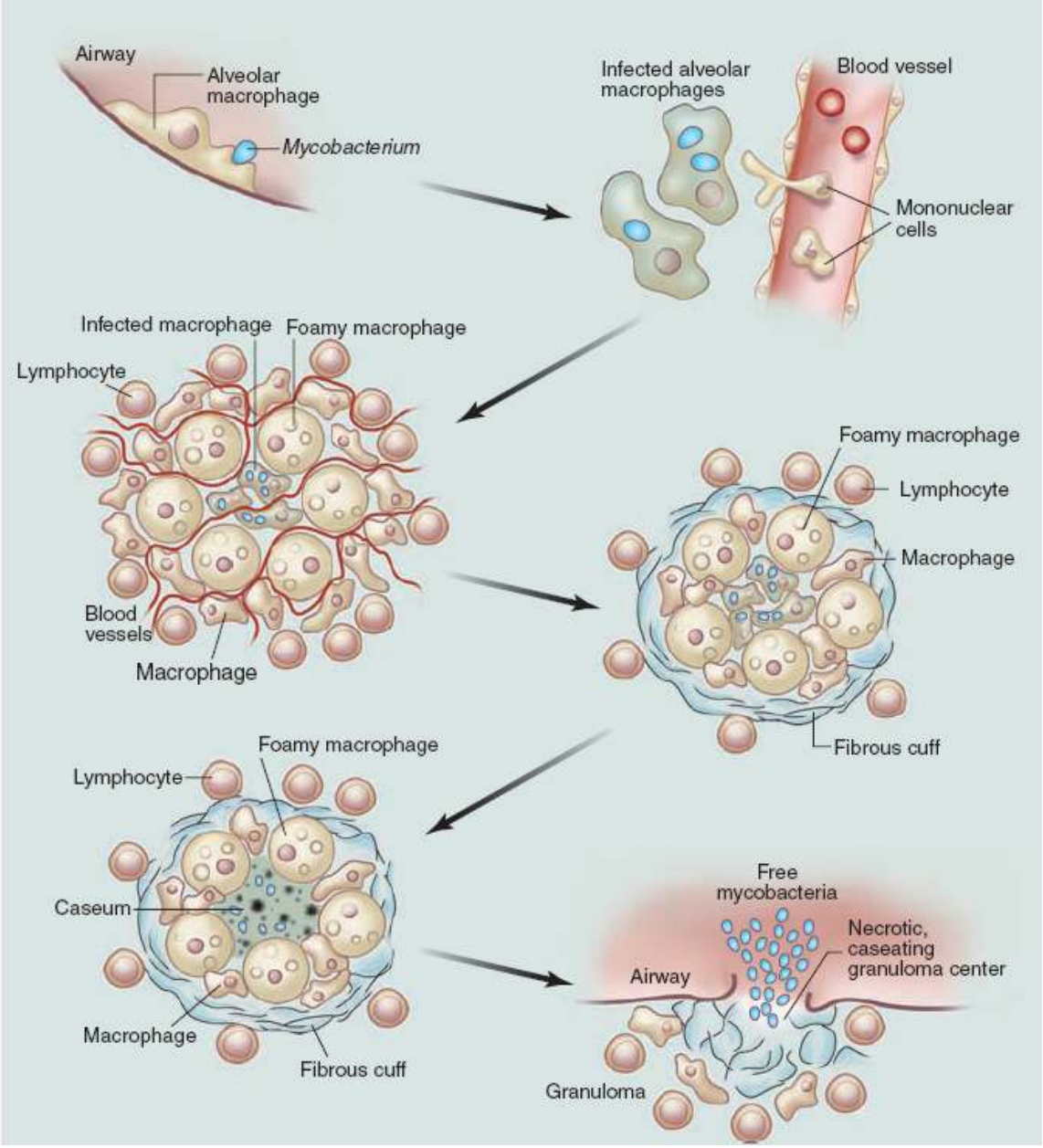


Figure 1.2 *M. tuberculosis* infection cycle. This figure depicts all stages of the *M. tuberculosis* infection cycle from infection of alveolar macrophages to the formation and then breakdown of the granuloma leading to a productive infection where the bacilli are expectorated. Figure adapted from Russell *et al*. (2010).

followed by treatment with only rifampicin and isoniazid for a further four months (Webb & Davies, 1999). Streptomycin is not currently part of the treatment regimen but is still widely considered a front line anti-TB drug, although when first used in the 1940's as a single antibiotic treatment it was found that within three months of treatment in 80% of cases resistant bacteria had evolved (Webb & Davies, 1999). Furthermore, those patients which test positive for TB by the skin test but do not display any clinical symptoms are required to undergo a six to nine month treatment with isoniazid alone (Russell, 2001).

However, patients infected with drug resistant strains of TB require a much longer treatment period with a larger cocktail of chemotherapeutics (Russell, 2001). Multidrug-resistant TB (MDR-TB) is diagnosed when the population of bacteria are resistant to both the front line drugs izoniazid and rifampicin. *M. tuberculosis* bacteria that are resistant to both front-line and second-line anti-TB drugs are becoming more widespread. This extensively drug-resistant TB (XDR-TB) is caused by bacteria that are resistant to izoniazid, rifampicin and also any fluoroquinolone treatments plus any of the intravenously administered anti-TB drugs including kanamycin, amikacin or capreomycin (WHO, 2010). As of March 2010 XDR-TB had been reported in 58 countries and territories (WHO, 2010).

Clearly the current treatment regimen is very long and coupled with the side effects of some of the drugs this leads to the problem of patient non-compliance. New drugs and treatment regimens are required to shorten treatment time and overcome the current problem of increasing drug resistance. New therapeutic compounds currently in the developmental pipeline include inhibitors of cell wall and protein synthesis, and more importantly inhibitors of ATP synthesis which will also hopefully target latent bacilli (figure 1.3). The fluoroquinolone compounds moxifloxacin and gatifloxacin (DNA gyrase inhibitors) are now in phase three development, compounds TMC207 and PA-824 are in phase two development whilst SQ-109 is under phase one development (figure 1.3) (Lalloo & Ambaram, 2010).

Furthermore, with one-third of the world's population latently infected with the TB bacillus the development of new treatments to stop the progression of latent infection into active disease would be an attractive and important goal (Stewart *et al.*, 2003).

1.1.2.3 Vaccination

The attenuated *M. bovis* bacillus Calmette-Guérin (BCG) strain is the only vaccine against *M. tuberculosis* currently in use and has variable efficacy around the world (Andersen & Doherty, 2005). The vaccine was produced in the 1920's by two French scientists who repeatedly passaged *M. bovis* for over 13 years to produce the attenuated BCG strain. It was first administered in 1921 in France and Belgium (Andersen & Doherty, 2005; Behr & Small, 1999), and after double-blind trials started in 1950 was subsequently introduced into the UK immunisation program in 1953 (Hart & Sutherland, 1977; Teo & Shingadia, 2006). In the UK, BCG vaccination is effective against pulmonary tuberculosis in children and provides approximately 70 to 80 percent protection, but it confers no protection in some third world regions; it is thought that genetic differences in populations and prior exposure to environmental mycobacteria that share antigens with *M. tuberculosis* results in a weaker immune response to BCG vaccination (Andersen & Doherty, 2005; Colditz *et al.*, 1994). Furthermore, BCG vaccination is largely ineffective at protecting against adult pulmonary disease in endemic TB areas (Colditz *et al.*, 1994). Various local seed lots of the initial BCG strain are used around the world to vaccinate the population either at birth or in childhood (Behr & Small, 1999). Routine BCG vaccination of school children began in the UK in 1953 but was stopped in 2005 as there was now only a low risk of TB infection in that population; however the vaccine is still given to those sections of the population considered to be at high risk from catching the disease (Teo & Shingadia, 2006).

Due to the variable efficacy of BCG in different populations around the world there is now an urgent need to develop new vaccines (Brennan, 2005; Fletcher & McShane, 2006; Fletcher *et al.*, 2009; Young & Dye, 2006). There are currently nine new TB vaccine candidates in clinical development and the hope is that three new vaccines will be in phase III trials by 2016 (Fletcher *et al.*, 2009). A new vaccine would ideally be effective in all ages, populations, races and be protective against all stages of the

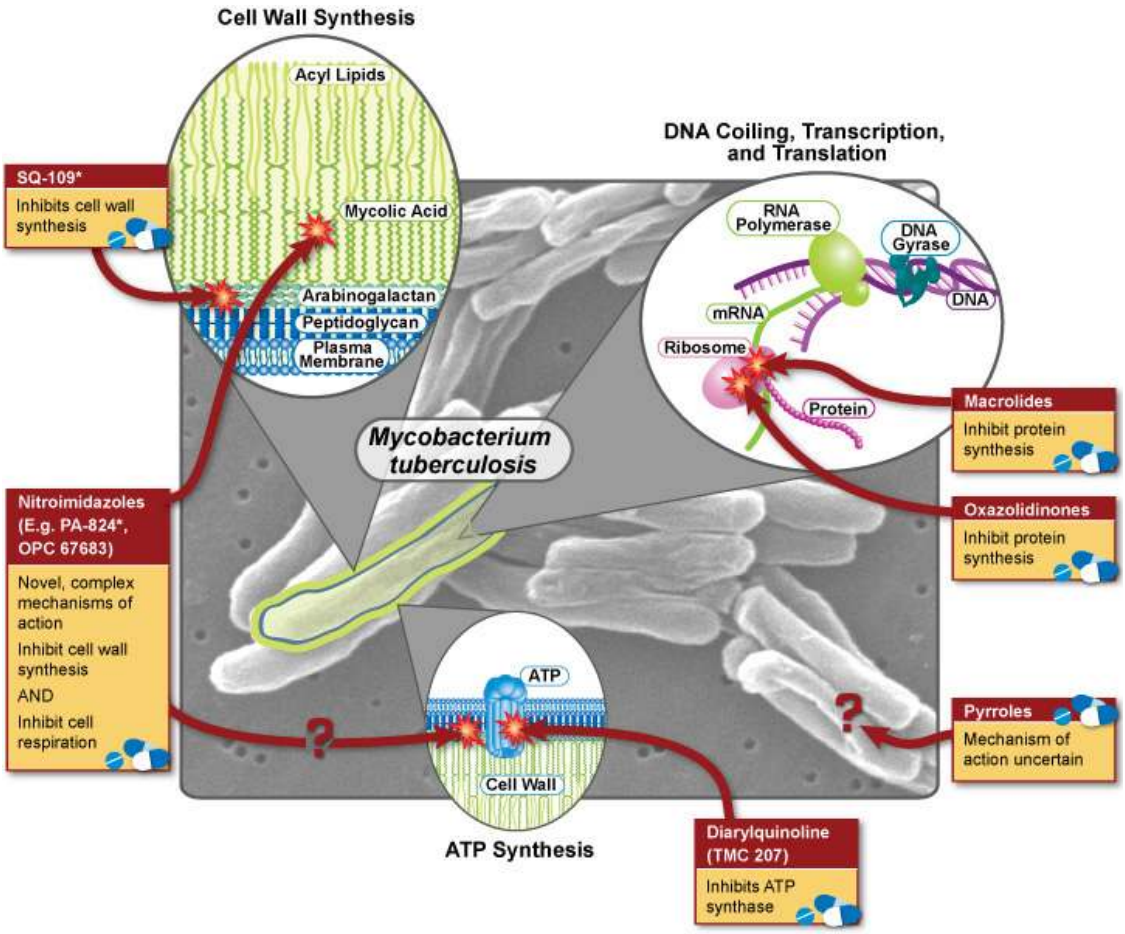


Figure 1.3 New TB drugs under development. Figure taken from the National Institute of Allergy and Infectious Diseases (NIAID) accessed at www.niaid.nih.gov/topics/tuberculosis/Understanding/WhatIsTB/ScientificIllustrations/pages/newtbdrugs.aspx. Compounds highlighted with asterisks are those which have been taken into development by NIAID.

infection including pulmonary, extra-pulmonary and latent stages of TB (Fletcher *et al.*, 2009). Besides such a prophylactic vaccine, a therapeutic vaccine is also required to deal with multidrug resistant TB. Vaccines under development fall under two main categories: live attenuated *M. tuberculosis* vaccines and subunit vaccines. Clinical trials of the first candidate subunit vaccine BCG-MVA85A (recombinant modified vaccinia virus Ankara (MVA) expressing antigen 85A) started in 2002 (McShane *et al.*, 2004; McShane *et al.*, 2005). It is hoped that boosting BCG vaccination with MVA85A could both enhance and prolong the antimycobacterial immunity (McShane *et al.*, 2004; McShane *et al.*, 2005). There are also several other versions of the Ag85A subunit vaccine in clinical trials some of which also include other mycobacterial antigens (Brennan *et al.*, 2007; Fletcher & McShane, 2006). Although subjected to tight regulatory and ethical requirements there are also five live *M. tuberculosis* vaccines currently in advanced pre-clinical development including an *M. tuberculosis phoP* deletion strain (Walker *et al.*, 2010).

Research into the function of key proteins is required to learn more about how *M. tuberculosis* causes disease and how it continues to be such a successful pathogen in order to produce novel treatments and find new drug targets to help eradicate the burden of this disease.

1.2 Protein Phosphorylation: A Universal Signal Transduction Pathway

Protein phosphorylation in signal transduction pathways is the most widespread post-translational mechanism (Ubersax & Ferrell, 2007). Phosphorylation has been implicated in nearly all basic cellular processes including metabolic regulation, transcription, cellular growth and movement, apoptosis, differentiation, cell cycle progression, intercellular communication, immunity, learning and memory (Manning *et al.*, 2002a; Manning *et al.*, 2002b; Ubersax & Ferrell, 2007). Phosphorylation is carried out by protein kinases.

Protein kinases catalyse the transfer of the γ-phosphate of ATP to a hydroxyl group on the side chain of specific amino acids in proteins whilst protein phosphatases function to

de-phosphorylate the protein or peptide through removal of the phosphate group. Phosphorylation of target proteins leads to transduction of the signal by creating conformational changes within the target protein, by creating opportunities for further intramolecular or electrostatic interactions and by further protein-protein interactions to activate or inhibit protein function (Ubersax & Ferrell, 2007).

Aberrant kinase activity is known to cause cancer as well as other diseases; therefore a large number of ATP competitive kinase inhibitors have been developed with the aim to treat these diseases (Lapenna & Giordano, 2009). Most small molecule kinase inhibitors are usually designed to compete with ATP for the ATP binding site thus functioning to inhibit autophosphorylation of the kinase, therefore keeping it in an inactive conformation. One hugely successful kinase inhibitor in the treatment of chronic myelogenous leukaemia (CML) and some other cancers is imanitib (Gleevec). Imanitib, developed by rational drug design, inhibits the activity of a tyrosine kinase called ABL and prevents its activity; in CML cells the tyrosine kinase enzyme is constitutively active (Lapenna & Giordano, 2009; Ren, 2005).

1.2.1 The protein kinase domain

The protein kinase domain (or catalytic domain) consists of a small N-terminal lobe (~ 85 amino acids) and a larger C-terminal lobe (~ 170 amino acids) (figure 1.4). The two lobes form a cleft where ATP can dock (figure 1.4) (Kornev *et al.*, 2006). Before kinases become active enzymes they have to convert from an inactive conformation into a catalytically competent form (Kornev *et al.*, 2006). This process is known as autophosphorylation of the protein kinase; one common mechanism of activation is through phosphorylation of the kinase activation loop on serine, threonine or tyrosine residues (figure 1.4) (Lochhead, 2009; Wang & Wu, 2002). Phosphorylation of these residues results in a conformational change in the activation loop rendering the kinase in a stabilised active conformation (Lochhead, 2009; Nolen *et al.*, 2004). There are three mechanisms by which the activation loop is phosphorylated: by an upstream kinase in a signalling cascade; by autophosphorylation in trans either where an active kinase trans-phosphorylates an inactive kinase or where two inactive molecules dimerise and

exchange activation segments; or by autophosphorylation in cis, an intramolecular event whereby kinases phosphorylate their own activation loops; much is still not known about this third mechanism of activation (Lochhead, 2009). Activated kinase domains play three main roles: binding and orientation of the ATP phosphate donor as a complex with a divalent cation (Mg^{2+} or Mn^{2+}); binding and orientation of the protein substrate; and transfer of the γ-phosphate from ATP to the acceptor residue of the protein substrate (serine, threonine or tyrosine) (Kornev *et al.*, 2006).

Archetypal protein kinases contain a conserved lysine-glutamic acid pair in their active site which forms critical contacts with the ATP (in figure 1.4 these residues are K72 and E91). The lysine residue binds to the α and ß phosphate groups of ATP and this residue is essential for enzyme activation (Young *et al.*, 2003). Two other important conserved motifs within kinase domains include the DFG motif and the HRD motif (where each letter represents the conserved amino acid). The aspartate residue in the DFG motif forms polar contacts with all three phosphate groups of ATP and is important for catalysis (figure 1.4) (Kornev *et al.*, 2006). The aspartate residue in the HRD motif helps to orientate the peptide substrate (Kornev *et al.*, 2006). Kinases with this motif are known as RD kinases. The arginine residue interacts with the phosphorylated activation loop and controls its fold finally resulting in activation of kinase activity (Kornev *et al.*, 2006).

1.2.2 *M. tuberculosis* signal transduction via protein phosphorylation

Reversible protein phosphorylation is a key mechanism used to transduce extracellular signals into cellular responses. Bacterial signal transduction systems can control many diverse processes including gene expression, adaptation to environmental stimuli, growth, development, stress responses, pathogenicity and host-pathogen interactions (Av-Gay & Deretic, 2005).

In *M. tuberculosis* signal transduction occurs via five main families of kinases and phosphatases (Chao *et al.*, 2010). The first family comprise the classic bacterial two-component systems involving histidine kinases and response regulators, the second family are the eukaryotic like serine/threonine protein kinases (STPKs), the third family

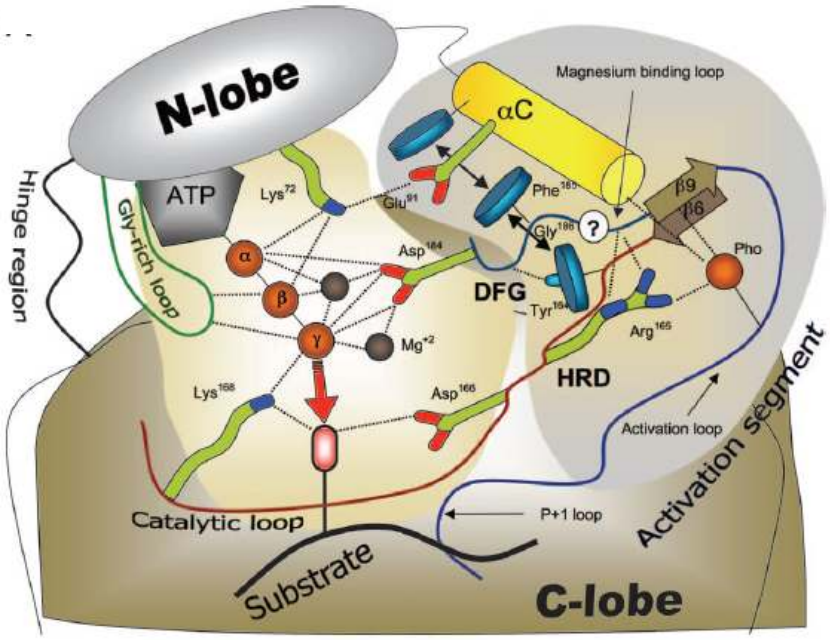


Figure 1.4 Diagram depicting the known interactions between the catalytic core of the protein kinase, ATP and a substrate protein. The amino acid numbering used in the diagram is for the eukaryotic cAMP-dependent protein kinase A. Image adapted from Kornev *et al*. (2006).

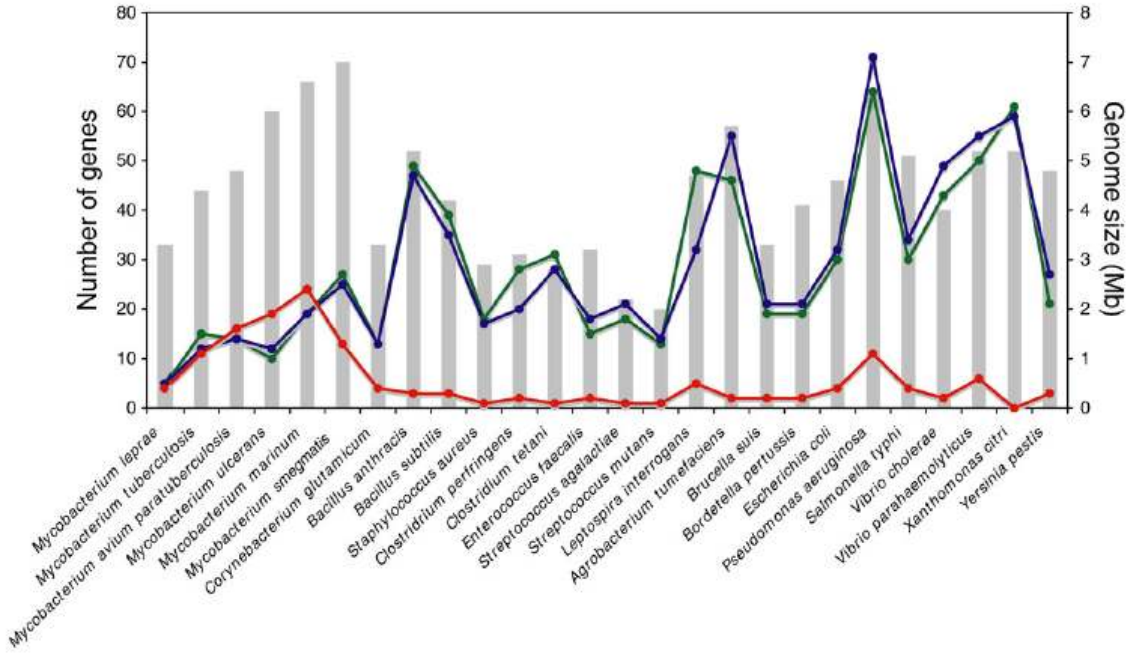


Figure 1.5 A comparison of signalling systems in bacteria. Figure adapted from Wehenkel *et al*. (2008). The genome sizes of different bacteria (in Mb) are shown by grey bars; the number of genes encoding histidine kinases are shown by blue dots and response regulators and shown by green dots; genes encoding STPKs are shown with red dots. The numbers of each of these systems were defined by homology searches.

comprises the single serine/threonine phosphatase, the fourth family comprises one protein tyrosine kinase, named PtkA, which has been recently identified and is a member of a newly identified class of tyrosine kinases (Bach *et al.*, 2009), and the fifth family comprises two secreted tyrosine phosphatases named PtpA and PtpB (Chao *et al.*, 2010; Koul *et al.*, 2000).

Signal transduction in *M. tuberculosis* has become a target for the development of novel therapeutics in the treatment of TB (Szekely *et al.*, 2008). Protein kinases and phosphatases allow reversible protein phosphorylation providing cells with a tool to sense and respond to both internal and external stimuli and thus represent attractive drug targets for new therapeutics.

1.2.2.1 Histidine kinases and response regulators: the classical bacterial two-component system

One major way prokaryotes sense their environment and regulate their gene expression profile is via the classical two-component signal transduction systems which were first described almost three decades ago (Av-Gay & Deretic, 2005; West & Stock, 2001). These systems are not present in humans (West & Stock, 2001). Two-component systems consist of a histidine sensor kinase protein and a response regulator protein. When the signal input domain of the sensor undergoes autophosphorylation a phosphate group from ATP is transferred from a conserved histidine residue on the sensor to a conserved aspartate or glutamate residue on the receiver domain to transduce the signal (West & Stock, 2001). Intriguingly, mycobacterial genomes contain relatively few histidine kinases and response regulators compared to other bacterial genomes of a similar size (figure 1.5) (Wehenkel *et al.*, 2008); although thirty genes encoding putative two-component system proteins have been identified in the *M. tuberculosis* genome (Cole *et al.*, 1998), only 11 systems paired in operons have been characterised. This is far fewer than in *E. coli* where over 30 pairs have been characterised (Mizuno, 1997). However, this can be explained by the fact that *M. tuberculosis* contains a relatively large number of STPKs; 11 eukaryotic-like STPKs and one tyrosine kinase are present in the genome meaning that phosphorylation can also occur on a serine, threonine or tyrosine residue. Therefore it is thought that in *M. tuberculosis* STPKs fulfil the roles of

the classical bacterial two-component systems (Av-Gay & Everett, 2000; Wehenkel *et al.*, 2008).

1.2.2.2 The 'eukaryotic-like' serine/threonine protein kinases of *M. tuberculosis*

STPKs involve the transfer of the γ-phosphate from ATP directly to a substrate and although originally thought to be present only in eukaryotes, these signalling domains have now been described in a number of bacteria (Bakal & Davies, 2000; Greenstein *et al.*, 2005). Eleven STPKs have been identified in the *M. tuberculosis* genome, named protein kinase A (PknA) to protein kinase L (PknL) (with the exception of PknC) (Cole *et al.*, 1998), nine of which are predicted to have a transmembrane-spanning region (Grundner *et al.*, 2005) (figure 1.6). Interestingly STPKs have been discovered in many species of pathogenic bacteria including *Listeria monocytogenes*, *Pseudomonas aeruginosa* and *Streptococcus pneumoniae* implicating STPKs in the regulation of virulence and pathogenesis of these organisms (Cozzone, 2005). For comparison, the human genome contains over 500 putative serine/threonine and tyrosine protein kinase genes (Manning *et al.*, 2002b), the yeast genome contains over 100, whilst only one is found in *E. coli* (Young *et al.*, 2003).

1.2.2.2.1 Structure, homology and activation of *M. tuberculosis* STPKs

Kinase domains of STPKs are approximately 250 amino acid residues in length and comprise 11 conserved subdomains which were initially described by Hanks *et al*. (Hanks *et al.*, 1988; Hanks, 1991; Hanks & Hunter, 1995). Like eukaryotic STPKs, *M. tuberculosis* kinase domains are activated through phosphorylation of specific residues in the activation loop; PknB, PknD, PknE and PknF show a conserved pattern of autophosphorylation in their activation loops (Boitel *et al.*, 2003; Duran *et al.*, 2005).

The structures of the kinase domains of PknB, PknD, PknE and PknG have been solved; the PknB, PknD and PknE kinase domains have been shown to be dimeric, which in the cell would presumably occur upon extracellular ligand binding to the sensor domain, and this dimerisation seemingly results in trans or cis autophosphorylation of the activation loop causing a conformational change leading to activation of the kinase (Gay *et al.*,

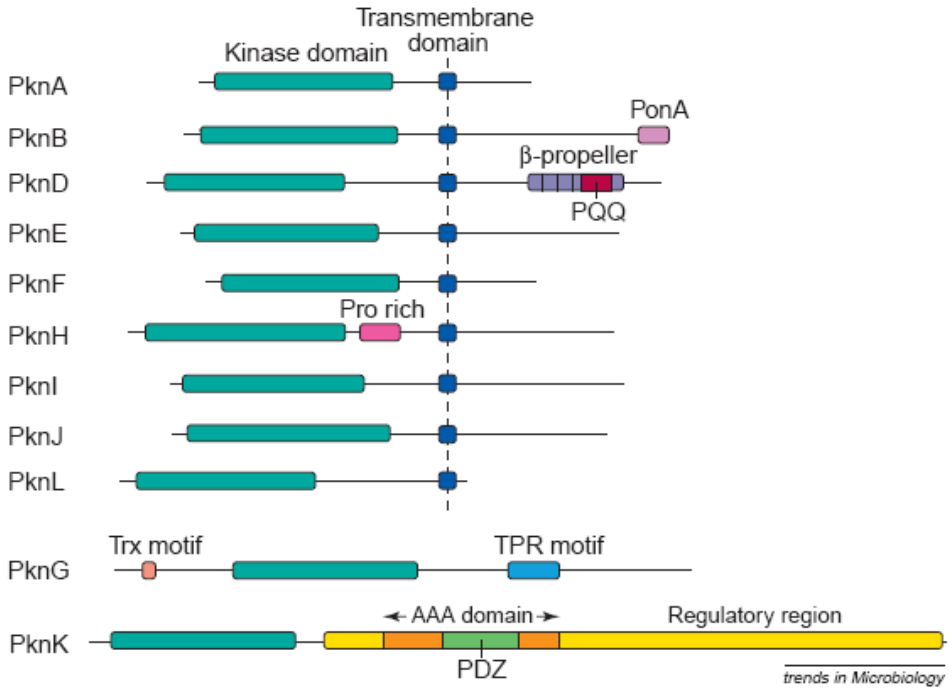
2006; Greenstein *et al.*, 2007a; Mieczkowski *et al.*, 2008; Ortiz-Lombardia *et al.*, 2003; Scherr *et al.*, 2007; Young *et al.*, 2003). This mechanism of activation is suggested to be conserved among other mycobacterial STPKs (Young *et al.*, 2003). However, PknG is unique among mycobacterial kinases as the critical arginine residue in the kinase domain is lacking (refer to section 1.2.1) (Av-Gay & Everett, 2000; Chao *et al.*, 2010). Therefore PknG is a non-RD kinase which suggests the absence of autophosphorylation in the activation loop; this was shown by O'Hare and Scherr (O'Hare *et al.*, 2008; Scherr *et al.*, 2009). PknG is therefore presumed to be activated via a different mechanism which involves the thioredoxin domain which senses the redox status of the cell (figure 1.6) (Av-Gay & Everett, 2000).

In *M. leprae*, where massive gene decay has occurred, only four of the *M. tuberculosis* STPKs are present, namely PknA, PknB, PknG and PknL, highlighting the probable essential roles of these proteins for bacterial viability (Cole *et al.*, 2001). Furthermore, all 11 *M. tuberculosis* STPKs are present in the *M. bovis* BCG vaccine strains (Behr *et al.*, 1999). Upon multiple sequence comparison of the *M. tuberculosis* STPKs with other prokaryotic kinases and eukaryotic kinases it was found that all the *M. tuberculosis* kinases clustered with other prokaryotic kinases with the exception of PknG and PknK which were most similar to the eukaryotic STPKs (Av-Gay & Everett, 2000). PknG and PknK are the two kinases with no transmembrane-spanning region and are therefore predicted to be soluble.

1.2.2.2.2 *M. tuberculosis* kinases as novel drug targets

The inhibition of essential signalling pathways in *M. tuberculosis* is an attractive prospect to yield novel classes of drug targets (Hegymegi-Barakonyi *et al.*, 2008). There has been much work focussed into designing inhibitors against the essential *M. tuberculosis* kinases PknA and PknB (Szekely *et al.*, 2008; Wehenkel *et al.*, 2006), although to date no kinase inhibitor has been discovered which has a high potency against *M. tuberculosis* despite appearing to be an effective inhibitor in *in vitro* assays (K. Lougheed, unpublished data) (Magnet *et al.*, 2010).

a. Serine/Threonine protein kinases



b. Serine/Threonine protein phosphatase

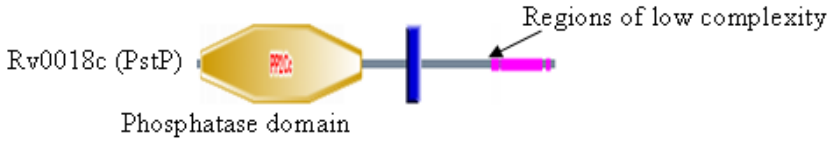


Figure 1.6 Structural analyses of *M. tuberculosis* serine/threonine protein kinases and protein phosphatase. The figure shows domain structures of the *M. tuberculosis* STPKs and the single serine/threonine phosphatase. Figure a adapted from Av-Gay and Everett 2000. The SMART domain database was used to construct the image in figure b (Letunic *et al.*, 2006). PQQ = pyrroloquinoline quinine, TPR = tetratricopeptide, Trx = thioredoxin.

1.2.2.2.3 M. tuberculosis STPK function

M. tuberculosis STPKs have been implicated to have diverse regulatory functions; PknA and B (the two essential STPKs) have been suggested to have roles in cell morphology and shape (Dasgupta *et al.*, 2006; Kang *et al.*, 2005). A recent study showed that InhA, involved in mycolic acid biosynthesis and the primary target of isoniazid, was phosphorylated *in vitro* by PknA, PknB, PknH and to a lesser extent by PknF demonstrating that at least *in vitro* there are overlapping phosphorylation motifs of the *M. tuberculosis* kinases (Khan *et al.*, 2010). PknA has also been shown to phosphorylate mtFabH (beta-ketoacyl-acyl carrier protein (ACP) synthase III) which is also involved in mycolic acid synthesis (Veyron-Churlet *et al.*, 2009) and FtsZ *in vitro* (Thakur & Chakraborti, 2006). PknA and PknB both phosphorylate the forkhead-associated domain-containing (FHA) protein Rv0019c (Gupta *et al.*, 2009; Sureka *et al.*, 2010) and EmbR *in vitro*, which is essential for the regulation of the *embCAB* operon encoding cell wall arabinosyltransferases (Sharma *et al.*, 2006); EmbR is also phosphorylated *in vitro* by PknJ (Jang *et al.*, 2010) and PknH (Molle *et al.*, 2003)). PknJ also phosphorylates pyruvate kinase A, an enzyme that catalyses the last step of glycolysis (Arora *et al.*, 2010). Furthermore, PknB phosphorylates PBPA *in vitro*, a penicillin-binding protein required for cell division (Dasgupta *et al.*, 2006) and both SigH and its cognate anti-sigma factor RshA *in vitro* and *in vivo* (Park *et al.*, 2008). SigH, which is negatively regulated by RshA, responds to oxidative, nitrosative and heat stress and is required for virulence in animal infection models (Park *et al.*, 2008).

PknD phosphorylates an FHA domain of Rv1747 *in vitro* (Grundner *et al.*, 2005) and Rv0516c, a putative anti-anti-sigma factor of unknown function *in vitro* and *in vivo* (Greenstein *et al.*, 2007b). A 2010 study by Vanzembergh and colleagues showed that PknD contributed to the survival of *M. tuberculosis* in phosphate poor conditions suggesting that PknD may play a role in detecting environmental phosphate (Vanzembergh *et al.*). In an earlier 2006 study MmpL7 was proposed to be a potential substrate for this kinase; PknD was shown to phosphorylate this protein which is required for the virulence of *M. tuberculosis* due to its role in PDIM translocation to the cell wall (Perez *et al.*, 2006).

PknE was shown to be involved in the nitric oxide stress response and apoptosis of *M. tuberculosis* in a human macrophage model of infection (Jayakumar *et al.*, 2008) whilst PknI has been implicated in controlling the slow growth of *M. tuberculosis* during infection and a mutant strain displayed a hypervirulence phenotype in SCID mice (Gopalaswamy *et al.*, 2009).

The phosphorylation of Rv1827 by PknB or PknG abrogates binding of Rv1827 to three proteins which are all involved in α-ketoglutarate metabolism (Nott *et al.*, 2009). Using pull-down assays and mass spectrometry it was found that PknG phosphorylates Rv1827 at threonine 21, adjacent to the residue phosphorylated by PknB (T22), and these two phosphorylation events are mutually exclusive (O'Hare *et al.*, 2008). Loss of Rv1827 phosphorylation leads to increased glutamate levels in the cell which is intriguing given the finding that a *pknG* deletion mutant led to a threefold higher accumulation of glutamine and glutamate in the cell compared with wild type (Cowley *et al.*, 2004). In the related actinobacterium *Corynebacterium glutamicum*, the homologue of Rv1827, Odh1, has been shown to be a regulator of a key enzyme of the tricarboxylic acid cycle, originally thought to be 2-oxoglutarate dehydrogenase (ODH) but now known to be 2-hydroxy-3-oxoadipate synthase (de Carvalho *et al*., 2010), and was shown to be phosphorylated by PknG (Niebisch *et al.*, 2006). In addition, an earlier paper published in 2004 suggested PknG promotes mycobacterial survival within macrophages by preventing phagosome-lysosome fusion; PknG was shown to be secreted in macrophage phagosomes and the inactivation of PknG resulted in mycobacterial cell death (Walburger *et al.*, 2004). It has also been shown that PknB, D, E and F can phosphorylate Rv1827 but with 5, 32 and 2 times lower that PknB activity again demonstrating that more than one STPK can phosphorylate the same substrate, at least *in vitro* (Villarino *et al.*, 2005).

PknL is predicted to have a role in transcriptional regulation and cell division (Lakshminarayan *et al.*, 2008) and phosphorylates the regulatory protein Rv2175c (Canova *et al.*, 2008). PknH has been suggested to play a role in regulating cell growth; a *pknH* deletion mutant actually resulted in a higher bacillary load in mouse organs compared to wild type (Papavinasasundaram *et al.*, 2005). Furthermore, PknK has been

implicated in a similar role where it is thought to be involved in early infection events in macrophages; a *ΔpknK* mutant showed increased survival during a persistent mouse infection thus implying a role in slowing down the growth of mycobacteria. The mutant was also more resistant to low pH, oxidative and hypoxic stress (Malhotra *et al.*, 2010).

Clearly, the signalling network of the *M. tuberculosis* STPKs is highly complex and much more work is needed to be performed in order to fully elucidate the inter-relationships and function of these important *M. tuberculosis* signalling proteins.

1.3 The forkhead-associated domain

Forkhead-associated (FHA) domain-containing proteins are phosphorylation-dependent proteins, conserved from bacteria to man, that specifically recognise phosphothreonine (pT) epitopes, distinguishing them from phosphoserine epitopes, and were initially identified within fork head transcription factors in 1995 (Hammet *et al.*, 2003; Hofmann & Bucher, 1995; Mahajan *et al.*, 2008). FHA domains are present in proteins with diverse functions including signal transduction, transcription factors, protein transport, cell cycle regulation and are particularly prevalent in DNA damage and repair genes (Mahajan *et al.*, 2008); furthermore in eukaryotes at least, FHA domains have been shown to be essential in the transduction of the phospho-signal (Durocher & Jackson, 2002). To date, FHA domains have been identified in more than 2000 prokaryotic and eukaryotic proteins (Mahajan *et al.*, 2008).

FHA domains form 11 stranded ß-sandwiches and are between 95 and 150 amino acid residues in size (Durocher, 2003; Hammet *et al.*, 2003; Mahajan *et al.*, 2008). The structure of the Rad53 FHA-1 domain from *Saccharomyces cerevisiae* was solved in complex with a phosphopeptide and is shown in figure 1.7 (Durocher & Jackson, 2002). Rad53 is a kinase which contains two FHA domains. Highly conserved residues within FHA domains are involved in direct contact with the phosphopeptide, either by contacting the peptide backbone or by forming a network of hydrogen bonds with the pT residue on the epitope (Durocher & Jackson, 2002). In the Rad53 FHA domain-containing protein, there are six highly conserved residues in the FHA-1 domain, five of which are located around the peptide binding site. Residues arginine-70 and serine-85

bind directly to the pT residue itself (Durocher *et al.*, 2000). Consequently, mutation of the conserved serine 85 residue within one of the peptide binding loops abolishes the binding activity of the FHA domain to the pT residue.

1.3.1 FHA domain-containing proteins in *M. tuberculosis*

Six FHA domain containing proteins have been identified in *M. tuberculosis*, four of which lie in putative operons with serine/threonine protein kinases; they include *embR* (*Rv1267c*) which forms a putative operon with *pknH*, *fhaA* (*Rv0020c*) and *fhaB* (*Rv0019c*) which both lie in the same putative operon with *pknA* (*Rv0015c*) and *pknB* (*Rv0014c*), and the focus of this project *Rv1747* which is in a putative operon with *pknF*. A schematic of the six *M. tuberculosis* FHA domains are shown in figure 1.8. Using the CLUSTAL W protein alignment program (http://www.ebi.ac.uk/Tools/clustalw2/index .html) I found that Rv3360 and Rv1747 FHA-1 are very similar sharing 62 % sequence identity and 78 % sequence similarity (figure 1.9). Intriguingly Rv3360 contains an alanine residue at a critical position where all the other FHA domains contain a serine residue necessary for the interaction of the FHA domain with phosphopeptides suggesting that the function of the FHA domain in Rv3360 may be abrogated (figure 1.9). FHA-1 and FHA-2 from Rv1747 share 26.2 % sequence identity. Of the *M. tuberculosis* FHA domain-containing proteins, only the EmbR structure has been solved in complex with a phosphopeptide; it contains the 11 stranded ß-sandwich seen in eukaryotic FHA domains (Alderwick *et al.*, 2006). Kinase activation loop phosphorylation renders the protein kinase in an active conformation to receive substrates. In the case of FHA domain-containing proteins it is thought that the peptide binding pocket of the FHA domain recognises and binds to one of the phosphorylated activation loop threonine residues on the kinase and then this delivers another part of the protein to the active site to be phosphorylated (Villarino *et al.*, 2005).

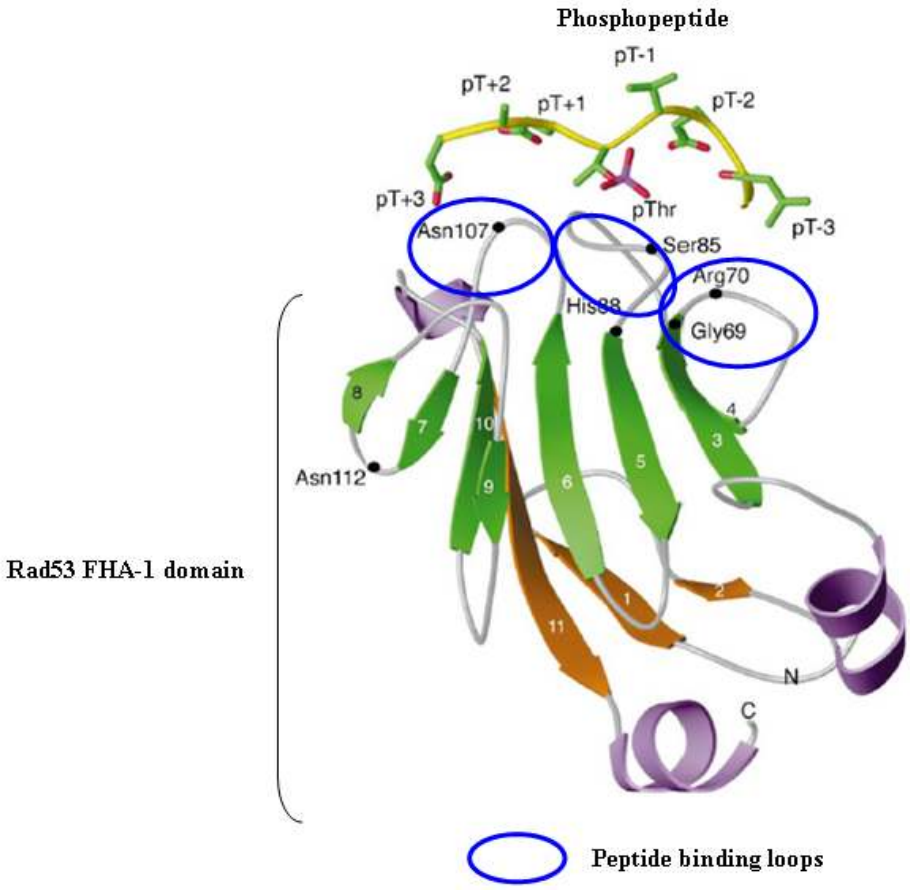


Figure 1.7 *S. cerevisiae* RAD53 FHA-1 domain in complex with a complex with a phosphothreonine-containing peptide. Figure adapted from Durocher and Jackson, (2002).

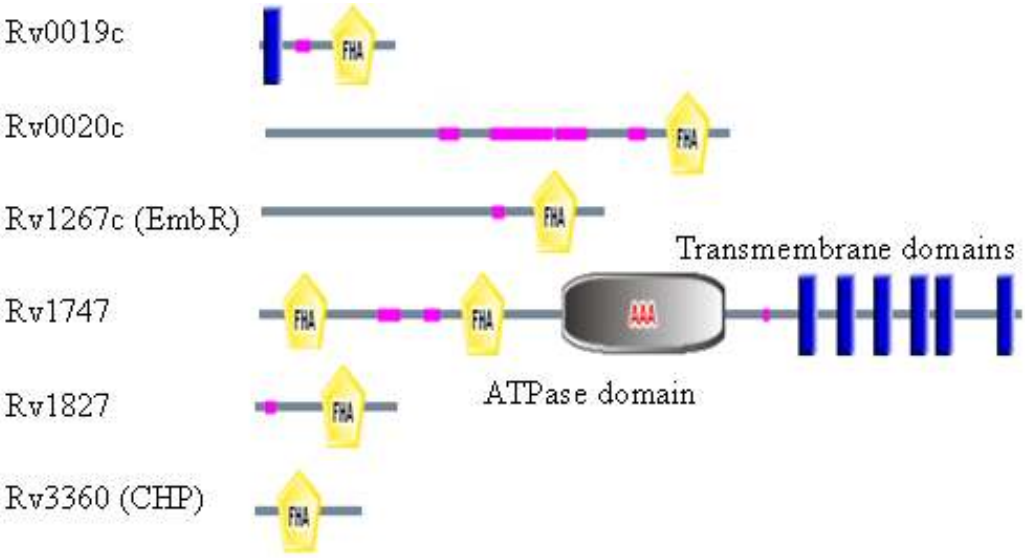


Figure 1.8 Structural analyses of the *M. tuberculosis* FHA domain-containing proteins. The SMART domain database was used to construct images in the figure (http://smart.embl-heidelberg.de/smart/set_mode.cgi?GENOMIC=1) (Letunic *et al.*, 2006). CHP = conserved hypothetical protein. Areas in pink denote regions of low complexity.

a. *M. tuberculosis* FHA domain-containing proteins alignment

```
Rv1747_FHA2     ----------------------------EAGNLATSMMKILRPGRLTG-----------
Rv1827          ----------------MTDMNPDIEKDQTSDEVTVETTSVFRADFLSELDAPAQAGT-E
Rv1747_FHA1     ----------------------------VPMSQPAAPPVLTVRYEGSER-----------
Rv3360          ----------------------------MSRP-HPPVLTVRSDRSQQ-----------
Rv1267c         AYYLSDRQSDALGAYRRVKTTLADDLGIDPGPTLRALNERILRQQPLDAKKSAKTTAAGT
Rv0020c         GATTYGRQDYGGGADYTRYTESPRVPGYAPQGGGYAEPAGRDYDYGQSGAPDYGQPAPGG
Rv0019c         LQLTRAGFLMLLWVFIWSVLRILKTDIYAPTGAVMMRRGLALRGTLLGAR----------

Rv1747_FHA2     -------ELPPG---------------------AVRIGRANDNDIVIPEVLASRHHATLV
Rv1827          SAVSGVEGLPPGSALLVVKRGPNAGSRFLLDQAITSAGRHPDSDIFLDDVTVSRRHAEFR
Rv1747_FHA1     -------TFAAG---------------------HDVVVGRDLRADVRVAHPLISRAHLLLR
Rv3360          -------CFAAG---------------------RDVVVGSDLRADMRVAHPLIARAHLLLR
Rv1267c         VTVLDQRTMASGQQAVAYLHDIASGRGYPLQAAATRIGRLHDNDIVLDSANVSRHHAVIV
Rv0020c         YSGYGQGGYGSAGTSVTLQLDDGSGRTYQLREGSNIIGRGQDAQFRLPDTGVSRRHLEIR
Rv0019c         -------QRRHAARYLVVTEGALTGARITLSEQPVLIGRADDSTLVLTDDYASTRHARLS
                              .                          *      . :     : *  :

Rv1747_FHA2     PTPGGTEIRDNRSINGTFVNGARVD-AALLHDGDVVTIGNID---LVFADGTLARREEN-
Rv1827          LENNEFNVVDVGSLNGTYVNREPVD-SAVLANGDEVQIGKFR---LVFLTGPKQGEDDGS
Rv1747_FHA1     FDQGRWVAIDNGSLNGLYLNNRRVP-VVDIYDAQRVHIGNPDGPALDFEVGRHRGSAGRP
Rv3360          FDRGNWIAIDNDSQSGMFVDGQRVS-EVDIYDGLTINIGKPTGPWITFEVGHHQGIIGRL
Rv1267c         DTGTNYVINDLRSSNGVHVQHERIRSAVTLNDGDHIRICDHEFT-FQISAGTHGGT----
Rv0020c         WDGQVALLADLNSTNGTTVNNAPVQ-EWQLADGDVIRLGHSEIIVRMH------------
Rv0019c         MRGSEWYVEDLGSTNGTYLDRAKVTTAVRVPIGTPVRIGKTAIELRP-------------
                        *  * .*  ::   :     :  .  : : .
```

b. Cladogram of the evolutionary relationships between the FHA domains

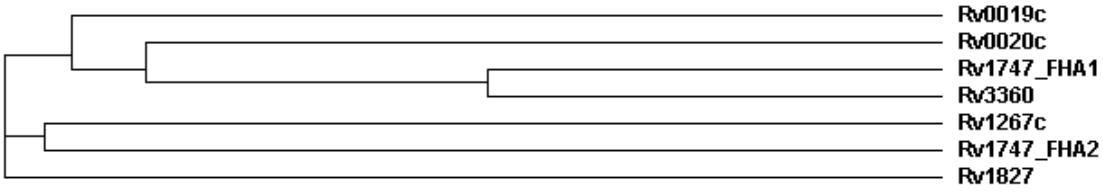


Figure 1.9 Alignment of the *M. tuberculosis* FHA domains Figure a is a sequence alignment and figure b is a cladogram of the evolutionary relationships between the FHA domains. Sequences were aligned using the Clustal W programme accessed at http://www.ebi.ac.uk/Tools/clustalw2/. Red residues = small amino acids, blue = acidic residues, magenta = basic residues and green = hydroxyl + amine + basic residues. * means that the residues or nucleotides in that column are identical in all sequences in the alignment. : means that conserved substitutions have been observed, according to the colour of the residues and "." means that semi-conserved substitutions are observed. The blue arrow highlights the serine residue critical for pT epitope binding which is not conserved in Rv3360.

1.4 PknF and Rv1747

This project focuses on PknF and Rv1747, an adjacent gene in the same putative operon coding for an ABC transporter (figure 1.10). PknF is a transmembrane STPK that contains an intracellular kinase domain and an extracellular domain presumed to be responsible for signal recognition (figure 1.10). Rv1747 contains two FHA modules, a feature unique to the six FHA domain-containing proteins of *M. tuberculosis* (figure 1.8).

1.4.1 ATP-binding cassette transporters

ATP-binding cassette (ABC) transporters form one of the largest families of transmembrane proteins thought to be present in all species from microbes to humans (Dassa, 2003).

ABC transporters bind and hydrolyse ATP providing energy for uptake or export of substrates across cell membranes and can be divided into three main groups: importers, exporters and those ABC proteins which do not appear to have a transport function but are involved in DNA repair and mRNA translation (Davidson *et al.*, 2008). Functions of ABC transporters include the uptake of nutrients into the cells and the export of virulence factors, toxins, proteins, lipids, and hydrophobic drugs (Davidson *et al.*, 2008). More diverse functions include roles in cell division, signal transduction, sporulation, nodulation, bacterial pathogenesis and antigen presentation (Davidson *et al.*, 2008; Higgins & Linton, 2003); cell survival is dependent on the transport of substrates across cell membranes.

ABC transporters are integral membrane proteins and the archetypal transporter comprises two hydrophobic transmembrane domains (TMDs) and two cytoplasmic nucleotide binding domains (NBDs), also known as ATPase domains. The two transmembrane domains span the membrane multiple times; in a typical transporter there are six membrane spanning segments per domain, a total of 12 per transporter (Higgins & Linton, 2003). The core ABC protein domain is approximately 215 amino acids in size (Higgins *et al.*, 1986); the NBD amino acid sequence contains three major

conserved motifs: the commonly found $Walker_A$ and $Walker_B$ motifs and a specific linker peptide signature motif located upstream of the $Walker_B$ motif (Saurin *et al.*, 1999; Schneider & Hunke, 1998; Walker *et al.*, 1982).

In humans, many ABC transporters are associated with disease; for example, mutations in ABC transporters are associated with genetic diseases including Cystic Fibrosis, Tangier disease and Macular Dystrophy; ABC transporters are also behind the problem of multidrug resistance in cancer therapy (Higgins, 2001; Linton & Higgins, 1998). Each year, thousands of cancer deaths are attributed to the overexpression of ABC transporters which results in resistance to chemotherapeutic drugs (Higgins & Linton, 2003).

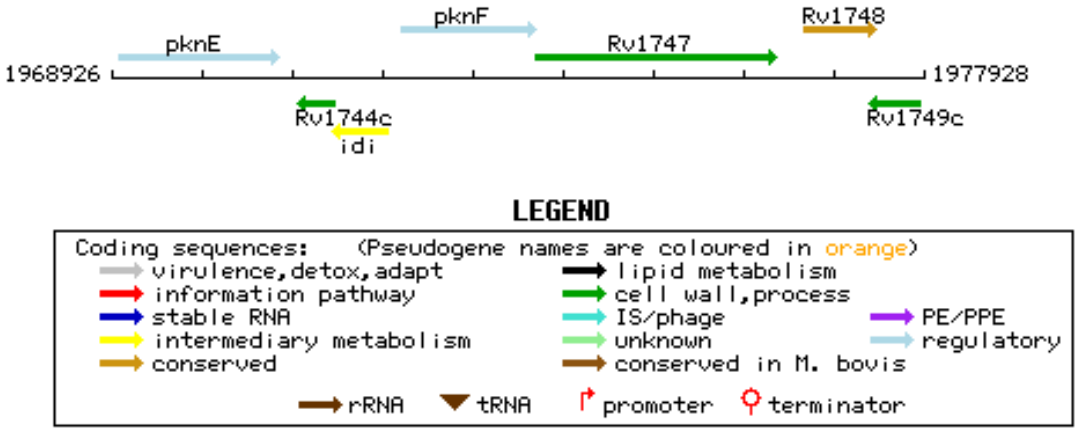
In bacteria, many ABC transporters are associated with being able to survive in different environments or are virulence determinants; with around 70 ABC transporters, approximately 5% of the *E. coli* genome is encoded by these transport molecules (Linton & Higgins, 1998). The number of ABC transporters encoded by different bacterial species varies widely and it is thought that this number is correlated with the diversity of environments that a particular species can adapt to live in i.e. bacteria that live in varied environments have more transporters than those which have a specific growth niche (Higgins & Linton, 2003).

Typical structures of a bacterial importer and exporter are shown in figure 1.11. Bacterial ABC importers are formed from four polypeptide chains, that are often separately encoded (Saurin *et al.*, 1999), and require an external binding protein which functions to deliver the substrate to the transporter (Dawson *et al.*, 2007). In contrast, bacterial exporters are produced as one polypeptide where a single gene usually encodes both the transmembrane domain and a NBD domain (Saurin *et al.*, 1999).

1.4.1.1 ABC transporters in *M. tuberculosis*

Approximately 2.5% of the *M. tuberculosis* genome comprises genes encoding putative ABC transporters; a total of 37 ABC transporters have been identified in *M. tuberculosis*, at least 26 of which appear to be complete i.e. contain at least one NBD

a. The *pknF* and *Rv1747* genes and surrounding genomic region



b. SMART domain structure of PknF

c. SMART domain structure of Rv1747

Figure 1.10 The *pknF* and *Rv1747* genes plus the domain architecture of the two proteins. The *pknF* and *Rv1747* genes are depicted in figure a. The marks on the chromosome are at 1000 bp intervals. Figures b and c depict the SMART domain structure of PknF and Rv1747 respectively. Regions shown in blue are the transmembrane domains. Figure b shows the kinase domain of PknF. Regions shown in pink denotes low complexity regions. Figure c shows the two FHA domains (yellow) and the ATPase domain of Rv1747 (Letunic *et al.*, 2006).

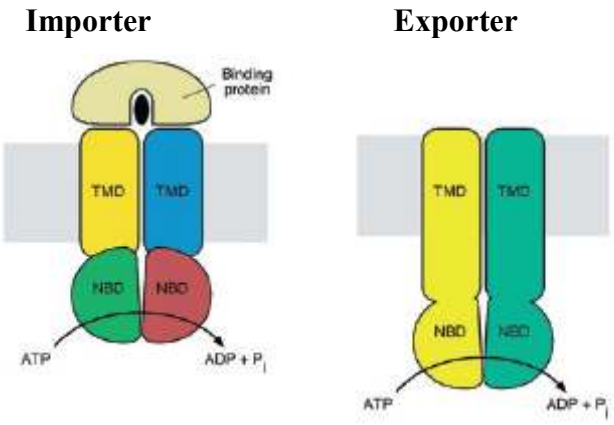


Figure 1.11 Typical structures of bacterial ABC transporters. This figure shows the domain structure of an archetypal bacterial importer and exporter. TMD = transmembrane domain. NBD = nucleotide binding domain. Image adapted from Dawson *et al*. (2007).

and one TMD. Sixteen of these transporters have been categorised as importers and 21 are presumed to be exporters (Braibant *et al.*, 2000).

The BacA (Rv1819c) ABC transporter is a recently characterised *M. tuberculosis* protein (Domenech *et al.*, 2009). A *ΔbacA* mutant strain showed identical growth in the lungs and spleens of mice over 205 days in an aerosol model of infection. However, in a time-to-death experiment the mice infected with the mutant strain survived significantly longer indicating a role for this transporter in maintaining a persistent infection in the host; this phenotype has been observed in *bacA* mutants of other alphaproteobacteria although the precise substrate of the *M. tuberculosis* BacA transporter remained elusive (Domenech *et al.*, 2009).

Rv1280c-Rv1283c encodes an oligopeptide importer which includes a substrate binding protein OppA (Rv1280c) (Dasgupta *et al.*, 2010). OppA was shown to bind the tripeptide glutathione and during a macrophage infection with an *ΔRv1281c* mutant there was a decrease in apoptosis and cytokine release from the macrophages (Dasgupta *et al.*, 2010). In 2008, it was shown that the ProXVWZ (Rv3759c-Rv3756c) ABC importer allowed *M. tuberculosis* to obtain betaine from human macrophages to help initiate mycobacterial growth; a *proXVWZ* mutant was impaired for survival and growth in macrophages (Price *et al.*, 2008). Together, these studies highlight the important role of ABC transporters in the virulence and persistence of *M. tuberculosis* infection.

M. tuberculosis ABC transporters have also been implicated in drug discovery; Rv1218c is an efflux pump and an *Rv1218c* mutant was found to be more susceptible to a variety of classes of compounds compared with wild type H37Rv suggesting that this ABC transporter effluxes the compounds from the cell (Balganesh *et al.*, 2010).

1.4.1.1.1 The ABC transporter Rv1747

The presence of fused nucleotide binding and transmembrane domains is a strong indicator of an ABC exporter (Davidson *et al.*, 2008; Dawson *et al.*, 2007). Based on its amino acid sequence Rv1747 belongs to the G subfamily of ABC transporters; this family consists of half-transporters which oligomerise to form the functional transporter

and have a unique domain arrangement where the nucleotide binding domain is located at the N-terminus of the protein with respect to the transmembrane spanning domains (Dassa, 2003). Furthermore an external binding protein required for the function of importers has not been identified for Rv1747 (Davidson & Maloney, 2007; Saurin *et al.*, 1999). Therefore, taken together, from its amino acid sequence and the fact that it is encoded by a single gene, Rv1747 is predicted to function as an ABC exporter.

However, although it has been demonstrated that the function of Rv1747 is required for the growth of the bacterium in mice and macrophages (Curry *et al.*, 2005), the identity of the substrate has not yet been revealed. Numerous hypotheses have been raised as to what the nature of the substrate may be by comparing the Rv1747 ABC transporter to analogous systems in other bacteria which are also required for growth and survival within the host. Bacterial ABC exporters transport many different substances which include cell surface components such as lipopolysaccharides, lipids, proteins involved in pathogenesis including haemolysin, and peptides, drugs and siderophores (Dassa & Bouige, 2001); Rv1747 could export any one of these molecules which would make the function of the transporter necessary for growth *in vivo*. A further discussion of possible substrates of Rv1747 is provided in the Introduction of Chapter 4.

ABCdb is a bacterial and archaeal ABC transporter database (http://www-abcdb.biotoul.fr/). According to this database Rv1747 was discovered to be similar to the gram positive soil bacterium *Streptomyces coelicolor* protein Q9AJX1 sharing 32.6 % identity with Rv1747. This protein contains two FHA domains, one nucleotide binding domain and five transmembrane domains. Furthermore the *Streptomyces avermitilis* protein Q829D9 also shares 31.5 % identity with Rv1747 and also contains the same domain structure suggesting that these two *Streptomyces* proteins are direct orthologs of Rv1747. Both these *Streptomyces* proteins are predicted to function as ABC transporters with currently no known function. Of the other *M. tuberculosis* ABC transporters Rv1747 is most closely related to Rv1687c, an uncharacterised ABC transport protein (Braibant *et al.*, 2000).

In humans, ABC transporters of the G family to which Rv1747 belongs, such as ABCG1, G4 and G5/8, are involved in ATP-dependent translocation of steroids and

possibly other lipids whilst ABCG2 (the breast cancer resistance protein) has been identified as a multidrug transporter (Velamakanni *et al.*, 2007).

It has been suggested (http://tuberculist.epfl.ch/) that the substrate of the Rv1747 ABC transporter may be a lipooligosaccharide. A review of all the *M. tuberculosis* ABC transporters published in 2000 revealed that Rv1747 falls into a sub-class of transporters which have an unknown function (Braibant *et al.*, 2000). Rv1747 is similar to the White protein (ABCG family of transporters) from *Drosophila melanogaster* (17% homology) which is a permease necessary for the export of pigment precursors responsible for eye colour, and to NodI from *Rhizobium* strains (21% homology) which is a protein implicated in the nodulation process by export of a lipooligosaccharide (Braibant *et al.*, 2000). I have compared the homology with other known ABC transporters from other organisms, and in addition to the two proteins mentioned above, an NCBI Blast of the Rv1747 protein revealed similarities with a cobalt transport family protein found in bacteria, eukaryotes and archaea and with an ABC group A subfamily which functions to transport a variety of lipid compounds suggesting that the Rv1747 transporter could be involved in any one of these functions in *M. tuberculosis*.

1.4.2 Previous studies on the PknF-Rv1747 system

Previous studies have investigated the interaction between PknF and Rv1747 and the possible substrates of Rv1747. A deletion mutant of *Rv1747* was shown to be attenuated in a mouse intravenous injection model of *M. tuberculosis* infection where the bacterial load of the mutant was 10-fold lower than that of the wild type in both lungs and spleen (Curry *et al.*, 2005). Furthermore, the Rv1747 strain was attenuated for growth in both naïve macrophages and dendritic cells (Curry *et al.*, 2005); in contrast the mutant grew as well as the wild type *in vitro* highlighting that this gene does not cause a general growth defect but is required for the full virulence of *M. tuberculosis* during infection thus warranting further studies to determine why this protein is important. A deletion mutant of *pknF* was also generated and a diagram showing the two gene deletions and the respective complementing strains is shown in figure 1.12.

Autophosphorylation of PknF occurs on threonine residues T173 and T175 of the activation loop (Duran *et al.*, 2005) and both FHA domains of Rv1747 seem to be required for the interaction of PknF and Rv1747 (Curry *et al.*, 2005; Molle *et al.*, 2004). Yeast-2-hybrid analysis demonstrated interaction between PknF and Rv1747. This interaction was abolished in a kinase null mutant of PknF and by mutation of the presumed activation loop of PknF and in one of the FHA domains of Rv1747 (Curry *et al.*, 2005). *In vitro* phosphorylation studies reported that the interaction was also stopped when one of the activation loop threonines known to be involved in PknF autophosphorylation was mutated (Molle *et al.*, 2004). Additionally, in the Molle *et al.* study it was also shown that the nucleotide binding domain of Rv1747 possessed ATPase activity suggesting that it was able to hydrolyse ATP to provide energy to transport a substrate across the *M. tuberculosis* membrane. Furthermore mutation of the serine residue in each of the FHA domains, which are essential for binding to the phosphothreonine residue, resulted in a reduced level of phosphorylation of Rv1747, thus highlighting the importance of the serine residue in these domains to mediate phosphorylation by PknF (Molle *et al.*, 2004).

In vitro assays have shown that PknF also phosphorylates another FHA domain containing protein, Rv0020c (Grundner *et al.*, 2005). Additionally, Molle *et al*. performed *in vitro* assays to demonstrate that PknF phosphorylates KasA, KasB (Molle *et al.*, 2006a) and the GroEL1 chaperone (Canova *et al.*, 2009). Furthermore, in 2009 the same group showed that PknF phosphorylates mtFabH to negatively regulate the FASII cycle involved in mycolic acid synthesis (Veyron-Churlet *et al.*, 2009). In addition, PknF was shown to phosphorylate InhA *in vitro*, but not as effectively as PknA, PknB or PknH (Khan *et al.*, 2010). Domains of Rv1747 have also been shown, at least *in vitro*, to be phosphorylated by PknB, PknD, PknE and PknF (Grundner *et al.*, 2005). Clearly therefore, there is a complex web of phosphorylation-dependent interactions.

A 2005 study reported that PknF negatively regulates glucose uptake; using antisense RNA to reduce the levels of PknF protein in *M. tuberculosis*, Deol *et al*. showed

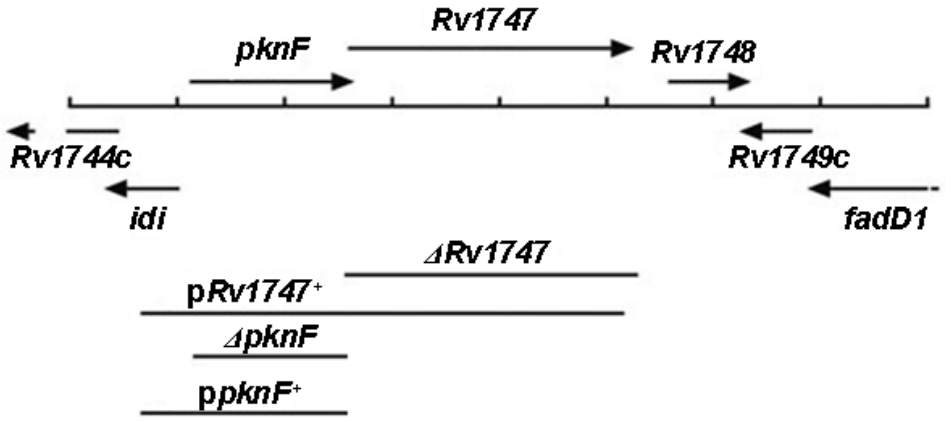


Figure 1.12 Diagram of the genomic region of *M. tuberculosis* containing the *pknF* and *Rv1747* genes. The figure shows the extent of the *pknF* and *Rv1747* deletions and the complementing plasmids p*pknF*$^{+}$ and p*Rv1747*$^{+}$. The *pknF* deletion was designed as an in frame deletion strain (R. Whalan, Personal Communication). The *Rv1747* complementing plasmid included a copy of *pknF*. The marks on the chromosome are at 1,000 bp intervals. Figure adapted from Curry *et al*. (2005) and R. Whalan (Personal Communication).

increased glucose uptake 16-fold higher compared with wild type and the antisense strain also grew 1.5 fold faster (Deol *et al.*, 2005). Additionally, the *M. tuberculosis* bacilli exhibited unusual cell morphology; the cells were shorter compared to wild type and the ends of the cells exhibited bulbous structures. This was interpreted as indicating that PknF may be regulating glucose uptake through phosphorylation of an ABC transporter (Deol *et al.*, 2005) although not presumably Rv1747 given this has a sequence characteristic of an exporter (figure 1.11) (Saurin *et al.*, 1999). The overexpression in *M. smegmatis* of the homologue of PknF resulted in irregular cell shape with bulbous like structures on the ends of the cells similar to the cell morphology observed after antisense expression (Deol *et al.*, 2005; Gopalaswamy *et al.*, 2007). Furthermore, *M. smegmatis* overexpressing the supposed PknF homologue (MSMEG3677) demonstrated reduced sliding motility and biofilm formation which was suggested could be due to alterations in glycopeptidolipid composition (Gopalaswamy *et al.*, 2007); glycopeptidolipids are only present among some non-tuberculous mycobacteria and are not found in *M. tuberculosis* (Billman-Jacobe, 2004). PknF belongs to the Clade III group of kinases along with PknI and PknJ (Narayan *et al.*, 2007). Using a comparative genomic approach Narayan and colleagues showed that whilst PknF has orthologs in all sequenced *M. tuberculosis* strains (H37Rv, CDC1551, F11 and *M.tb* C), *M. bovis*, *Mycobacterium avium* ssp. *paratuberculosis* and *Mycobacterium marinum*, when looking at the co-localisation of genes, i.e. looking for an Rv1747 ortholog, they could not assign an ortholog in *M. smegmatis* or *M. leprae* since none of the Clade III members from these species showed conservation of synteny (Narayan *et al.*, 2007).

It is therefore of interest to investigate why Rv1747 is required for virulence, how PknF is controlling Rv1747, and to determine what molecules are transported by Rv1747 in order to understand how the *pknF-Rv1747* signal transduction system functions in *M. tuberculosis*.

1.5 Project Hypotheses and Aims

This project focuses on investigating the function of PknF and Rv1747 in *M. tuberculosis*. The overarching hypothesis is that serine-threonine protein kinases and

their substrates are required for normal growth of *M. tuberculosis in vivo* and as such represent potential targets for drug intervention.

1.5.1 Hypothesis

Rv1747 is an ABC exporter, whose activity is controlled by phosphorylation mediated by PknF, that exports a cellular component essential for virulence and whose expression is controlled by stress to the bacterium.

1.5.2 Aims

- To characterise the *pknF-Rv1747* operon:
    - To investigate the *pknF* stimulus,
    - To determine whether transcription of the *pknF-Rv1747* genes is affected by external stresses to the bacterium. For example it is known that transcription of the *pknE* gene is affected by nitric oxide (Jayakumar *et al.*, 2008).
    - To determine whether the *pknF* and *Rv1747* mutant strains are more susceptible to any drug or stress reagent.
- To investigate the nature of the Rv1747 substrate:
    - To determine if mutation of *pknF* or *Rv1747* affects the transcription of other genes, which may give a clue to what is being transported by Rv1747.
    - To determine if there are changes in cellular composition of the *pknF* and *Rv1747* mutants: lipid, protein, small molecule or metabolite.
- To determine how PknF is controlling Rv1747 function:
    - To analyse the phosphorylation dependent interactions between the two proteins.
        - To determine the sites where PknF phosphorylates Rv1747 to give information about how the proteins function together.
- To characterise the *pknF-Rv1747* signalling system and analyse the importance of both genes for the virulence of *M. tuberculosis*:
    - To use *M. tuberculosis* infection models to assess the importance of different protein domains.

CHAPTER 2. MATERIALS AND METHODS

Composition of all commonly used media, buffers and reagents are listed in Appendix I.

2.1 Bacterial Strains and Plasmids

Strains and plasmids used for cloning and expression of mycobacterial proteins are shown in tables 2.1 and 2.2. The mycobacterial strain used was *Mycobacterium tuberculosis* H37Rv (Steenken *et al.*, 1934). All *Escherichia coli* strains were grown on L. agar and in L. broth supplemented with appropriate antibiotics (table 2.3) overnight at 37 ^{o}C (with shaking for liquid cultures (250 rpm)). *M. tuberculosis* strains were grown on 7H11 agar plus 10 % OADC (Becton and Dickinson) containing appropriate supplements and antibiotics (table 2.3). Liquid cultures were grown in 50 ml Falcon tubes (Corning) or 1000 ml polycarbonate roller bottles (Nalgene) containing Dubos medium supplemented with 0.2 % v/v glycerol and 4 % v/v Dubos medium albumin (Becton and Dickinson) or Sauton's medium plus or minus 0.05 % Polysorbate (Tween) 80 (Sigma) containing appropriate supplements (table 2.3) at 37 ^{o}C in a Bellco roll-in incubator (at 2 rpm).

Strain	Description	Source
***Escherichia coli* Strains**		
E.coli One Shot TOP10 cells	*E. coli* cloning strain: F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G F^-: does not carry the F plasmid *mcrA*, *mcrBC*, *mrr*: mutations allowing methylated DNA not to be recognized as foreign. Important when cloning genomic/methlyated DNA. *lac*ZΔM15: partial deletion of *lacZ* that allows for blue/white screening of recombinant clones by alpha complementation *rec*A1: recombination deficient, increasing insert stability. *ara*D139, *gal*U, *gal*K: blocks ability to catabolise arabinose, leucine, galactose. *rps*L: streptomycin resistant *end*A1: endonuclease deficient. The mutation increases plasmid yield and quality *nup*G: Mutation for the transport of nucleosides	Invitrogen
E.coli DH5α Library Efficiency cells	*E. coli* cloning strain: F^-80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (r_k^-, m_k^-) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^- *deo*R: allows constitutive expression of nucleoside catabolism genes *hsd*R17: prevents methlyated DNA being recognized as foreign, allows for efficient transformation, *sup*E44: tRNA glutamine suppressor *gyr*A96: DNA gyrase mutant. *rel*A1: mutation to allow RNA synthesis to occur in the absence of protein synthesis. In wild type limiting amino acid concentrations results in the shutdown of RNA synthesis.	Invitrogen
E.coli BL21 Star (DE3) cells	Protein expression strain.*E. coli* F^- *ompT hsd*SB(r_B^-, m_B^-)*gal dcm rne*131(DE3) *ompT*: outer membrane protease deficient designed to maximize the amount of recombinant protein λ(DE3): bacteriophage λ carries the gene for T7 RNA polymerase which is integrated into the host genome and controlled by the *lavUV5* promoter. Contains *lacI* repressor: IPTG binds the lac repressor and de-represses the promoter; *dcm*: abolishes methylation at specific sequences to enable cleavage with certain restriction enzymes. *rne*131: RnaseE mutant, protects against mRNA degradation.	Invitrogen
E. coli BL21 star (DE3) pRep4-*groESL*	Protein expression strain. *E. coli* BL21 star (DE3) plus pRep4-*groESL* plasmid pRep4-*groESL*: Plasmid which expresses GroES and GroEL to increase protein solubility and yield. Kanamycin resistant.	V.Molle
SoloPack Gold Supercompetent cells	*E. coli* cloning strain used for site directed mutagenesis: Tet^r Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 *thi*-1 *recA*1 *gyr*A96 *rel*A1 *lac Hte* [F′ *pro*AB *lacIqZ*ΔM15 Tn10 (Tetr) Cam^r]. *lacI*q*ZΔM15* gene: allows blue-white screening for recombinant plasmids. Hte: High transformation efficiency. Tet^r: Tetracycline resistant, Cam^r: Chloramphenicol resistant	Stratagene

Table 2.1. *E. coli* Strains used in this study

Plasmid	Description	Source
pCR4Blunt-TOPO	General cloning vector; Kan^r, Amp^R	Invitrogen
pEJ414	pMV306 derivative containing a promoterless *E. coli lacZ* reporter gene, Kan^R	Papavinasasundaram *et al.*, 2001
pGEX-6P-1	Replicating protein expression vector. N-terminal GST-Tag, *tac* promoter, *lacI* repressor, Amp^R	GE Healthcare
pBS-Int	Suicide vector containing integrase, Amp^R	Springer, 2001
P2Nil	Suicide gene delivery vector, *oriE,* Kan^R	Hinds *et al.*, 1999
pRW69	p2Nil containing *Rv1747*, Hyg^R	Curry *et al.*, 2005
pKP186	Integrase negative derivative of the integrating vector pMV306, Kan^R	K.G. Papavinasasundaram
pRW_76	*Rv1747* complementing plasmid. pKP186 derivative containing 609 bp *Rv1745*, p*knF* and *Rv1747*, Kan^R Hyg^R	R. Whalan
pRW_95	*PknF* complementing plasmid. pKP186 derivative containing 609 bp *Rv1745*, *PknF* and 20 bp *Rv1747*, Kan^R	R. Whalan
pRW_70	pRW_76 with S47A mutation in *Rv1747*	R. Whalan
pRW_71	pRW_76 with S248A mutation in *Rv1747*	R. Whalan
pVS_01	pEJ414 containing *pknF* promoter region, Kan^R	This study
pVS_02	pGEX-6P-1 containing $pknF^{1-292}$, Amp^R	This study
pVS_03	pGEX-6P-1 containing $Rv1747^{1-559}$, Amp^R	This study
pVS_04	pGEX-6P-1 containing $FHA1^{1-120}$, Amp^R	This study
pVS_05	pGEX-6P-1 containing $FHA2^{202-310}$, Amp^R	This study
pVS_06	pGEX-6P-1 containing $FHA1^{1-120}$ S47A, Amp^R	This study
pVS_07	pGEX-6P-1 containing $Rv1747^{1-559}$ T150A, Amp^R	This study
pVS_08	pGex6P-1 containing $Rv1747^{1-559}$ T150D, Amp^R	This study
pVS_09	pGex6P-1 containing $Rv1747^{1-559}$ T208A, Amp^R	This study
pVS_10	pGex6P-1 containing $Rv1747^{1-559}$ T208D, Amp^R	This study
pVS_11	pGex6P-1 containing $Rv1747^{1-559}$ T150A/ T208A, Amp^R	This study
pVS_12	pGex6P1 containing $Rv1747^{1-559}$ T150D/ T208D, Amp^R	This study
pVS_13	pRW76 containing $Rv1747^{1-559}$ T150A mutation	This study
pVS_14	pRW76 containing $Rv1747^{1-559}$ T150D mutation	This study
pVS_15	pRW76 containing $Rv1747^{1-559}$ T208A mutation	This study
pVS_16	pRW76 containing $Rv1747^{1-559}$ T208D mutation	This study
pVS_17	pRW76 containing $Rv1747^{1-559}$ T150A/T208A mutations	This study
pVS_18	pRW76 containing $Rv1747^{1-559}$ T150D/T208D mutations	This study
pVS_19	pGEX-6P-1 containing $pknF^{1-292}$ K41A, Amp^R	This study

Table 2.2. Plasmids used and constructed in this study

Antibiotic/Supplement	**Concentration used in *E. coli* (µg/ml)**	**Concentration used in *M. tuberculosis* (µg/ml)**
Kanamycin	50	25
Ampicillin	100	N/A
Gentamycin	20	15
Hygromycin	250	50
Chloramphenicol	34	N/A
X-gal	200	50
Sucrose	N/A	20

Table 2.3 Antibiotics and supplements added to growth media.

2.2 Recombinant DNA Techniques

2.2.1 Design of oligonucleotides

Primer sequences used in this study are listed in Appendix II. Oligonucleotides were designed according to the parameters set by the program PRIMER DESIGN FOR WINDOWS (Scientific and Educational Software). They were typically 16 to 25 base pairs in length and were checked for the presence of hairpins which could inhibit PCR. Primers were synthesised by Eurogentec or Sigma and were supplied desalted at 40 nmol scale synthesis. Alternatively if they were to be used for site directed mutagenesis they were supplied RP-HPLC purified.

2.2.2 Polymerase chain reaction (PCR)

PCR facilitates amplification of template DNA using cycles of changing temperature. Template DNA is heated to force strand dissociation, followed by annealing of specific oligonucleotides to the DNA strand by lowering the temperature, and completed by an extension step which uses DNA polymerase to incorporate dNTPs and produce a complementary strand of DNA. For amplifications where it was necessary to use a proof reading enzyme eg. cloning, reactions were performed in 50 µl volumes in 0.2 ml thermo tubes containing 1 X *Pfx* Amplification buffer, 1 X Enhancer buffer (lowers the DNA melting temperature), 0.3 mM dNTPs, 0.3 µM of each oligonucleotide, 1 mM $MgSO_4$, 80 ng *M. tb* DNA and 2.5 U *Pfx* enzyme (Invitrogen). DNA amplification was performed in an MWG Primus thermocycler using a touchdown cycle. Touchdown cycles help reduce non-specific background by starting with an annealing temperature a few degrees above the T_m of the primers and then lowering by one degree every two cycles. This increases the likelihood of primers binding their specific template DNA. A further 20 to 25 cycles are then employed to amplify the product. Cycles consisted of an initial 95 ^oC denaturation step for five minutes followed by the touchdown cycles with an annealing temperature from ~ 65 to 56 ^oC and extension period at 72 ^oC for one minute per kilobase of product. 20 to 25 cycles were then performed with an annealing temperature of 56 ^oC and completed with a final extension time of eight minutes.

For DNA amplifications where the proof reading capacity of DNA polymerase was not necessary PCR's were performed using REDTaq Readymix (Sigma) in 25 µl volumes in 0.2 ml thermo tubes containing 1 X REDTaq buffer, 0.2 mM dNTPs, 0.3 µM of each oligonucleotide, 1.5 mM $MgCl_2$, 1.5 U Taq DNA polymerase, and one to five microlitres of Instagene DNA using the same touchdown cycle. HotStar*Taq* polymerase (Qiagen) was used for Reverse Transcription PCR following the same touchdown cycle and including a 15 minute initiation step at 95 ^{o}C. HotStarTaq is a modified *Taq* polymerase which is completely inactive at room temperature and requires a 15 minute step at 95 ^{o}C to activate the enzyme, thereby reducing non-specific amplification of products where primers have bound during PCR set up.

2.2.3 Site directed mutagenesis (SDM)

SDM oligonucleotides were designed with the desired mutation in the middle of each primer with approximately 20 nucleotides of correct sequence on each side and a melting temperature of approximately 78 ^{o}C. Primers were synthesised and RP-HPLC purified by Eurogentec or Sigma.

PCR amplifications were performed in 50 µl volumes in 0.2 ml thermo tubes containing 1 X *Pfx* Amplification buffer, 1 X Enhancer buffer, 0.3 mM dNTPs, 125 ng of each oligonucleotide, 1 mM $MgSO_4$, 100-300 ng plasmid DNA and 2.5 U Pfx (Invitrogen). DNA amplification was performed in an MWG Primus thermocycler. An initial denaturation step at 95 ^{o}C for one minute was followed by 18 to 25 cycles of 95 ^{o}C for 50 seconds, an annealing step of 60 ^{o}C for 50 seconds and a 68 ^{o}C extension temperature for two minutes per kb of plasmid and then completed by a final extension at 68 ^{o}C for seven minutes. Parental methylated DNA was then digested by addition of 1 µl *Dpn*I (Fast Digest, Fermentas) and incubated at 37 ^{o}C for 15 minutes. DNA was then precipitated as outlined in section 2.2.6. One to three microlitres of resulting plasmid DNA was then used for transformation into *E. coli* Supercompetent cells (Strategene) using the heat-shock method as outlined in section 2.2.9. Plasmid DNA was extracted from the resulting colonies (2.2.4) and was sequenced as in section 2.2.12.

2.2.4 Agarose gel electrophoresis

Agarose gel electrophoresis separates DNA fragments based on mass and charge. When a current is applied, the negatively charged DNA molecules migrate towards the positive pole. Smaller DNA fragments migrate through the gel faster than larger fragments. Ethidium bromide intercalates between DNA base pairs and is therefore used to visualise DNA.

Depending on expected product/s size 0.8 to 2 % agarose gels were cast using 1 X TBE. The solution was boiled to dissolve the agarose, 0.5 µg/ml ethidium bromide (Bio-Rad) was added and gels were cast with a comb. Once set, gels were immersed in 1 X TBE. 1 X loading buffer (30% glycerol, 0.25% bromophenol blue, 0.25% xylene cyanol) was added to each sample, 8 µl of a 1 kilobase DNA ladder (Hyperladder I, Bioline) was run alongside as a marker and voltage was applied at 100 V for 40 minutes or until DNA had sufficiently migrated. DNA fragments were visualised using a UV transilluminator (BioDoc-It Imaging System).

2.2.5 Plasmid DNA preparation

Plasmid DNA was purified from 5 ml of stationary phase *E. coli* cultures using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's guidelines. Briefly, bacteria were lysed under alkaline conditions to denature proteins and DNA, and SDS solubilized phospholipids and proteins of the cell membrane. Lysates were neutralised and adjusted to high salt conditions which precipitates denatured proteins, detergent, debris and chromosomal DNA whilst plasmid DNA renatures and stays in solution. Plasmid DNA was purified by adsorption onto a silica membrane, contaminants were washed away with an ethanol based solution then DNA was eluted in 50 µl dH_2O.

2.2.6 DNA fragment extraction and purification from agarose gels

DNA fragments were excised from the gel using a UV lightbox (at 302 nm) and scalpel. DNA was extracted and purified using the QIAquick Gel Extraction kit (Qiagen) following the manufacturer's guidelines. Agarose gel slices were solubilised in a binding buffer with optimal pH and salt concentrations for DNA to be adsorbed to the silica

membrane. DNA was then adsorbed to the membrane, contaminants were washed away using an ethanol based wash solution and then DNA was eluted in 30 µl dH_2O.

2.2.7 Ethanol precipitation of nucleic acids

DNA or RNA was precipitated from solution by addition of 0.1 volume 3M sodium acetate and 2.5 volumes cold 100 % ethanol, incubated at - 20 ^oC overnight and then centrifuged for 15 minutes in order to form a pellet. The pellet was washed with 70 % ethanol, left to dry at room temperature and then resuspended in dH_2O.

2.2.8 Restriction endonuclease digestion of DNA

Restriction digestion of DNA was performed in 15 µl reactions, for one hour at 37 ^oC or five minutes at 37 ^oC if using FastDigest Enzymes (Fermentas), containing the DNA and restriction enzyme/s in appropriate buffers. Vector digests were subsequently treated with 5 U Antarctic Phosphatase (NEB) for 20 minutes at 37 ^oC followed by a five minute incubation at 65 ^oC to deactivate the enzyme. This step removes the 5´ phosphate group from linearised vectors to prevent vector re-ligation thereby increasing cloning efficiency by reducing vector background.

2.2.9 Ligation of DNA

Ligations were performed in 10 µl reactions using a 1:7, 1:5 or 1:3 ratio of vector:insert DNA using T4 DNA Ligase and 1X T4 DNA Ligase buffer (Promega). Ligations were performed at room temperature for four hours or at 4 ^oC for 16 hours.

2.2.9.1 Ligation of DNA into pCR4Blunt-TOPO

PCR products were initially cloned into pCR4Blunt-TOPO vector. The vector is supplied in a linearised form with topoisomerse I bound to the 3´ ends. Topoisomerase I ligates the PCR product into the vector and then dissociates from the DNA. 2 µl purified PCR product was combined with 0.5 µl pCR4Blunt-TOPO and 0.5 µl salt solution (1.2 M NaCl), then incubated at room temperature for one hour.

2.2.10 Transformation of *E. coli*

Transformations were performed with chemically competent *E.coli* One Shot TOP10, *E.coli* DH5α and *E.coli* BL21 Star using a heat shock transformation procedure or with electrocompetent *E.coli* BL21 Star pRep4 cells (table 2.1). Competent cells were defrosted on ice then 100 to 200 ng of plasmid DNA was carefully mixed with the cells and incubated on ice for 15 to 30 minutes. For heat shock transformation, cells were incubated at 42 ^{o}C for 30 to 45 seconds to facilitate uptake of plasmid into the cells. For electrocompetent strains, cells were electroporated with 2.5 kV, 1000 ohms and 25 μF in a Gene Pulser chamber (Biorad). Finally, LB or SOC medium was added to the cells and incubated at 37 ^{o}C with shaking for one hour to allow expression of selective markers before being plated on selective LB agar.

2.2.11 Colony Screening and Glycerol Stocks

Colonies potentially containing the desired plasmid insert were screened by PCR as described in section 2.1. Plasmid DNA was subsequently extracted from clones which appeared correct and then checked by sequencing. Glycerol stocks of correct clones were made by combining 850 μl of overnight culture with 150 μl sterile glycerol and stored at -80 ^{o}C.

2.2.12 Determination of nucleic acid and protein concentration

The concentration of nucleic acids was determined using a Nanodrop ND-100 spectrophotometer, Version 3.1.2 (Labtech). 1.2 μl of sample was added to the spectrophotometer and the concentration readout was noted. To determine protein concentration the molecular weight and extinction coefficient of each protein was included in the calculation.

2.2.13 DNA Sequencing

Sequencing of PCR products and plasmids was performed by Geneservice Limited (Cambridge) containing 1 μg of plasmid DNA or 200 ng PCR product and 32 pmoles of each primer.

2.3 Transcriptional Analysis

2.3.1 Reverse Transcription PCR

Reverse transcription PCR was employed to assess the co-transcription of genes of interest using the Quantitect Reverse Transcription Kit (Qiagen) which uses a high efficiency reverse transcriptase enzyme capable of detecting low abundance transcripts. A reverse transcription reaction was set up containing 1 µg of DNA free RNA in 1 x Quantiscript RT buffer containing magnesium and dNTPs, RT primer mix and 1 µl Quantiscript reverse transcriptase. Reactions were incubated at 42 ^{o}C for 30 minutes to allow cDNA synthesis to occur. A negative control was also set up for each sample replacing the reverse transcriptase with 1 µl dH_2O. Samples were then incubated at 95 ^{o}C for three minutes to inactivate the reverse transcriptase enzyme. PCR was then performed using the cDNA template with HotStarTaq as outlined in section 2.1.

2.3.2 5´ Rapid Amplification of cDNA Ends (5´ RACE)

5´ RACE was performed to determine the transcriptional start site of genes of interest using the 5` RACE Kit (Invitrogen) following the manufacturer's guidelines. Three Gene Specific Primers (GSPs) were designed. GSP 1 is responsible for first strand cDNA synthesis and was designed to anneal at least 300 bp downstream from the mRNA 5´ end to allow for cDNA purification on a column. GSP 2 is responsible for the first PCR so was designed to anneal 3´ with respect to the cDNA. GSP 3 is used in a nested PCR to generate sufficient specific product for sequencing so was designed to anneal within the primary PCR sequence.

cDNA synthesis was performed using 5 µg of *M. tuberculosis* RNA and 2.5 pmoles GSP1. The reaction was incubated at 70 ^{o}C for 10 minutes to denature the RNA then chilled for one minute on ice. First strand cDNA synthesis was then performed by addition of 1 x PCR buffer (20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$), 10 mM DTT, 0.4 mM dNTPs and 200 U SuperScript II Reverse Transcriptase (Invitrogen) to the RNA and incubated at 42 ^{o}C for 50 minutes. cDNA synthesis was then stopped by

incubation at 70 ^{o}C for 15 minutes. After cooling to 37 ^{o}C 1 µl of RNase mix was added and incubated for 30 minutes to remove the original mRNA template.

cDNA was purified on a column to remove unincorporated primer and dNTPs. Briefly, samples were bound to the column under high salt binding conditions, washed three times with cold 1 x wash buffer then twice with cold 70 % ethanol. cDNA was eluted in 50 µl dH2O.

A poly-C tail was then added to the 3′ ends of the cDNA using Terminal deoxynucleotidyl Transferase (TdT) and dCTP. TdT tailing creates a binding site for the Abridged Anchor Primer which was used with GSP2 to allow direct amplification by PCR using *Taq* DNA polymerase. Once complete, a second nested PCR amplification was then performed with the Abridged Universal Anchor Primer and GSP3 in order to generate sufficient specific product for sequencing. PCR products were separated on an agarose gel, extracted, purified and sequenced.

2.3.3 Quantitative Real-Time PCR (qRT-PCR)

qRT-PCR uses a fluorescence based PCR reaction to allow detection of accumulating nucleic acids in real time. SYBR green I (Applied Biosystems), a fluorophore which intercalates double-stranded DNA, was used for qRT-PCR. The dye fluoresces when bound to double-stranded DNA enabling detection of DNA products as they accumulate during cycles of PCR.

cDNA was generated from 1 µg RNA (section 2.4.3) using the Quantitect reverse transcription kit (section 2.3.1). Due to assay sensitivity cDNA was diluted appropriately. DNA standards were generated using H37Rv genomic DNA. Primers were designed according to the parameters set in the Primer Express 3.0 program (Applied Biosystems). Reactions were performed in biological triplicate in 20 µl reactions with 2 x Fast SYBR green master mix (Applied Biosystems), 900 nM of each primer plus 5 µl DNA standard, cDNA or water (as a no template control) in MicroAmp fast optical 96-well reaction plates (0.1 ml) (Applied Biosystems). Absolute quantification qRT-PCR was performed as per manufacturer's guidelines on a 7500

Real-Time PCR System (Applied Biosystems) using the following conditions: a 95 ^{o}C step for 20 seconds to activate enzymes in the master mix followed by 40 cycles with a 95 ^{o}C denaturation step for three seconds and 60 ^{o}C annealing and extension steps for 30 seconds. A dissociation step was performed at the end of qRT-PCR. This step gives the melting temperature (T_m) of each PCR product to indicate if more than one product was amplified (i.e. to check the PCR was specific).

Data were analysed using the 7500 Fast System Software (Applied Biosystems). Quantities of cDNA within the samples were calculated from the cycle threshold (Ct) values, the cycle at which the fluorescence increased above the baseline level, using the appropriate standard curve. A reverse transcription negative control reaction was performed for each sample to control for any genomic DNA contamination. These values were subtracted from the reverse transcription positive values. All data was averaged and normalized to *sigA* expression to give the normalised transcription level of genes of interest.

2.4 Mycobacteria Specific Techniques

All *Mycobacterium tuberculosis* specific techniques were performed in Category Three Containment facilities using class I hoods.

2.4.1 Preparation and transformation of electrocompetent *M. tuberculosis*

Rolling cultures (100 ml) were grown to an OD_{600} of 1.0 before 0.1 volume of 2M glycine was added and incubated for a further 20 to 24 hours. Addition of glycine improves transformation efficiency of mycobacteria; glycine replaces alanine during peptidoglycan synthesis resulting in reduced cross-linking. Cells were harvested by centrifugation at 16, 000 rcf for 20 minutes. Cell pellets were resuspended in 100 ml 10 % glycerol then re-harvested. A total of three washes were required. The cell pellet was finally resuspended in 10 ml 10 % glycerol. Electrocompetent cells were stored at 4 ^{o}C or - 80 ^{o}C until use.

2.4.1.1 Electroporation of mycobacteria

400 µl of electrocompetent cells were added to an electroporation cuvette with the required amount of DNA (1 to 5 µl volume). Cuvettes were placed into the Gene Pulser chamber (Biorad) and electroporated with 2.5 kV, 1000 ohms and 25 µF. Cells were then added to 3.6 ml Dubos and recovered overnight at 37 ^{o}C. After recovery cells were diluted appropriately (usually 1/10 and 1/100 fold), plated on 7H11 agar plus appropriate supplements and selection agents and incubated at 37 ^{o}C for two to four weeks until colony growth was observed.

2.4.2 DNA isolation from *M. tuberculosis* on 7H11 agar plates and in liquid culture

Two methods were employed to extract DNA; when screening potential *M. tb* clones the Instagene Method (Biorad) was used whereas when high molecular weight good quality genomic DNA was needed either for PCR or whole genome sequencing a more stringent method was applied. Both methods are outlined below.

1. Instagene Preparation of DNA

A colony scraped from a plate or a pellet from 100 µl of mid-log culture was resuspended in 200 µl Instagene Matrix (Biorad) and incubated at 56 ^{o}C for 30 minutes. This step helps the matrix to bind to components in the preparation that may inhibit PCR. Samples were then boiled at 100 ^{o}C for eight minutes in a waterbath to both kill the *M.tb* and release genomic DNA. One to five microlitres of this preparation was then used in a colony screening PCR reaction as outlined in section 2.2.1.

2. Genomic DNA isolation from *M. tuberculosis* for whole genome sequencing

20 ml of an exponential phase culture was harvested by centrifugation at 1,700 rcf for 15 minutes and the pellet was resuspended in 1 ml TE buffer. Bacteria were heat killed by incubation at 80 ^{o}C for one hour. Cells were pelleted by centrifugation at 13, 000 rcf and resuspended in 1 ml resuspension buffer. Bacteria were pelleted and washed in 450 µl of resuspension buffer. 100 µl of 10 mg/ml lysozyme (Sigma) was added and incubated at 37 ^{o}C for two hours to facilitate cell lysis. 50 µl of 20 % SDS (Biorad) and 25 µl of 20

mg/ml Proteinase K (Sigma) was added to each sample and incubated at 55 ^{o}C for 40 minutes to denature membrane lipids and digest proteins. 250 µl of 4M NaCl and 160 µl of centrimide saline solution was then added and incubated at 65 ^{o}C for 10 minutes to provide a buffer solution for extraction of DNA. One ml of chloroform:isoamyl alcohol (24:1) was added to each sample then the aqueous phase containing the nucleic acids was extracted by vigorous vortexing followed by centrifugation. The aqueous phase was transferred to a new tube and the previous step was repeated. 0.7 volumes of isopropanol was added and mixed to precipitate the DNA. The genomic DNA spool was extracted, washed in 70 % ethanol, dried for 15 minutes at room temperature and then resuspended in 50 µl TE buffer. DNA was confirmed to be all high molecular weight and not degraded by separating 3 µl on a 0.8 % agaroase gel. 5µg of this DNA was then sent to GATC Biotech for whole genome sequencing using Illumina's Genome Analyser (Illumina).

2.4.3 RNA isolation from *M. tuberculosis* liquid cultures

Total RNA was isolated from 100 ml of mid-log phase rolling cultures using the Fast RNA Pro Blue kit (Qbiogene) as per the manufacturer's guidelines. Cultures were harvested by centrifugation and then resuspended in 2 ml RNAPro solution which stabilizes RNA and inactivates RNases during cell lysis to prevent degradation. The resuspension was transferred into two matrix tubes and ribolysed (Hybaid) on setting six for 40 seconds to release the cellular RNA, DNA and proteins. The tubes were centrifuged at 13, 000 rcf for 5 minutes at 4 ^{o}C then the upper phase was transferred to a new tube. 300 µl of chloroform was added, tubes were vortexed and then incubated for five minutes at room temperature before being centrifuged as previously. The upper aqueous phase containing the RNA was transferred to a new tube, being careful not to touch the interphase or organic phase containing proteins and other contaminants.

RNA was precipitated by addition of 500 µl of 100 % ethanol and incubated at - 20 ^{o}C overnight. The RNA was then pelleted by centrifugation and washed in 70 % ethanol (made with DEPC treated water). The pellet was allowed to air-dry and then resuspended in 100 µl DEPC H_2O containing 40 U RNase inhibitor (Promega).

Contaminating DNA was removed by DNase digestion with 2 U RNase free DNase (Promega) in 5 mM magnesium sulphate and 100 mM sodium acetate with 80 U RNase inhibitor and incubated at 37 ^{o}C for one hour. A further 2 U of DNase was then added and incubated for a further hour as before. The presence of contaminating DNA was then checked using 35 cycles of touchdown PCR (section 2.2.1).

Proteins and other contaminants were removed from RNA samples using the RNA cleanup kit (RNeasy kit, Qiagen) following the manufacturer's guidelines. Briefly, RNA samples were suspended in a buffer containing guanidinium thiocynate (to denature the enzymatic activity of RNases) and ethanol which creates optimal conditions for RNA to bind to the spin column. Samples were then applied to the spin column. Contaminants were washed away with an ethanol based wash solution and RNA was eluted twice in 30 µl of RNAse-free water. 200 ng of clean RNA was then run on a 2100 Bioanalyzer (Agilent Technologies) to assess integrity (R. Butler, NIMR).

2.4.4 Generation of *M. tuberculosis* cell free extract

Bacteria were isolated from exponential phase cultures (OD_{600} 0.6-0.8) by centrifugation at 1,700 rcf for 20 minutes, washed in PBS, and then cells were broken in a ribolyser (Hybaid) on setting 6.5 for 30 seconds with 150-212 µm glass beads (Sigma). Culture supernatants were then filtered through low binding 0.22 µm Durapore membrane filters (Ultrafree-MC, Millipore) by centrifugation at 13,000 rcf for 5 minutes. Alternatively, cultures were isolated by centrifugation and supernatants filtered twice through 0.22 µM filters.

2.4.4.1 Generation of *M. tuberculosis* protein extract for assessment of protein composition in culture supernatants

Exponential phase cultures (OD_{600} 0.6-0.8) were centrifuged as outlined in section 2.4.4 and then supernatants were filtered twice through 0.22 µM filters. Trichloracetic acid (TCA) was then used to precipitate the proteins from the *M. tuberculosis* culture supernatants; TCA treatment ensured each protein sample was as concentrated as possible. 10% TCA was added to each sample of culture supernatant in a 1:1 v/v ratio.

Samples were incubated on ice for 30 minutes to allow precipitation to occur. They were then centrifuged at 13,000 rcf for 15 minutes at 4 ^{o}C to pellet the proteins. Pellets were washed in 300 µl of cold acetone, centrifuged as previously and then the pellet was allowed to dry before resusupension in dH_2O. Half of each TCA precipitated sample was then digested with trypsin; this enzyme is a serine protease which cleaves proteins at the carboxyl side of arginine or lysine. 0.2 µg of trypsin (Sigma) was added to 100 µl of TCA precipitated protein sample and then incubated at 37 ^{o}C for 16 hours. TCA precipitated samples and trypsin digested samples were then analysed on a 12% SDS-PAGE gel as described in section 2.5.1.

2.4.5 *M. tuberculosis* RNA-DNA microarrays

Whole genome microarrays of *M. tuberculosis* (version 2) were obtained from by Dr. J. Hinds and Dr. K. Gould (St. George's, University of London). 7 µg of total RNA was labelled with 3 µg of random primers (Invitrogen) in the presence of 500 U SuperScript II Reverse transcriptase (Invitrogen), 1x first strand buffer, 10 mM DTT, 5 mM dA/G/TTP, 2 mM dCTP and 1.5 nmol Cy3 or Cy5 labelled dUTP. RNA and random primers were heated at 95 ^{o}C for five minutes in a final volume of 11 µl and snap cooled for two minutes. 500 U SuperScript II Reverse transcriptase (Invitrogen), 1x first strand buffer, 10 mM DTT, 5 mM dA/G/TTP, 2 mM dCTP and 1.5 nmol Cy3 or Cy5 labelled dUTP was then added to the RNA and random primer mix and incubated at 25 ^{o}C for 10 minutes, followed by 42 ^{o}C for 90 minutes to allow cDNA synthesis to occur.

Microarray slides were prehybridised in 3.5 x SSC, 0.1 % SDS and 10 mg/ml BSA (Sigma) for 20 minutes at 60 ^{o}C and then washed for one minute in water and one minute in isopropanol. Slides were then placed inside 50 ml Falcon tubes and centrifuged at 300 rcf for five minutes to dry the slides. Labelled samples were competitively hybridised against a single microarray using two Lifterslips (Thermo Scientific) to cover the array, combining the samples and then pipetting the sample under the Lifterslips allowing the solution to cover the microarray by capillary action. Microarray slides were placed in hybridisation chambers, submerged in water and incubated in a 65 ^{o}C oven for 16 to 24 hours. Slides were then washed vigorously for

two minutes in 1 x SSC, 0.05 % SDS followed by two washes in 0.06 x SSC for two minutes each.

Microarray slides were scanned with an Axon GenePix 4000A Microarray Scanner (Molecular Devices) using PMT voltages of 600-800 and then analyzed using GenePix Pro. Grids were fitted using Bluefuse software, and further analysis was performed using the GeneSpring 10 (Agilent) software program. Three biological replicates were performed for each condition and were carried out in duplicate (dye-swaps, to correct for dye-labelling bias). See appendix III for further details on microarray analysis.

2.4.6 Lipid extraction and two dimensional thin layer chromatography of *M. tuberculosis* cell wall lipids

Exponential phase cultures were labelled with 1 μCi/ml [^{14}C] acetate (~300mCi/mmol) for 24 hours. Cultures were then washed thrice in PBS; resulting pellets were incubated at 80 ^{o}C for one hour whilst supernatants were filtered twice through low binding 0.22 μm Durapore membrane filters (Ultrafree-MC, Millipore). Supernatants were then freeze dried using an E-C Modulyo (E-C. Apparatus Inc, USA) freeze-drier.

I performed all thin layer chromatography experiments at the University of Birmingham in Professor Del Besra's laboratory. To extract non-polar (apolar) lipids each sample was dissolved in 2 ml CH_3OH : 0.3% NaCl (100:10, v/v) and 1 ml of petroleum ether (60-80 ^{o}C), transferred to a glass vial (Tube A) and mixed on a rotator for 15 minutes before centrifugation at 2,000 rcf for five minutes. The upper layer was transferred to a separate tube (Tube B). One ml of petroleum ether was added to the lower layer (A) then mixed and centrifuged as previously. The upper layer was transferred to tube (B). The upper layers (non-polar lipids) were then dried in a Techne Sample concentrator at 50 ^{o}C.

To extract polar lipids 3 ml of $CHCl_3$:CH_3OH:NaCl (0.3 %) 90:100:30 was added to the bottom layer (Tube A), mixed as previously for 60 minutes on a rotator and centrifuged for five minutes at 2,000 rcf. Supernatants were transferred to separate tubes (Tube C). 750 μl of $CHCl_3$:CH_3OH:0.3% NaCl 50:100:40 was added to the pellet, mixed for 30

minutes and centrifuged as previously. The supernatant was added to tube C. This step was repeated once more. The remaining pellet contained the non-extractable lipids. 1.3 ml of $CHCl_3$ and 1.3 ml of 0.3% NaCl was added to the pooled sample (Tube C), mixed on the rotator for five minutes and centrifuged for three minutes at 2,500 rcf. The lower layer contained the polar lipids. This was removed into a separate tube and dried as for non-polar lipids. Each dried sample was resuspended in 200 µl $CHCl_3$: CH_3OH 2:1 and centrifuged at 2,000 rcf for five minutes. 5 µl was spotted onto the bottom of a scintillation vial, allowed to dry, then 5 ml scintillation fluid was added, vortexed, and the radioactivity measured in a scintillation counter. Two dimensional thin layer chromatography was then performed to separate the lipids by spotting 20,000 cpm of each sample onto a silica plate (Merk). The lipids were separated in five systems in two directions using different combinations of solvents (table 2.4). Apolar lipids were separated in systems A to D and polar lipids were analysed in systems D and E. The silica plates were allowed to dry fully in between each run. The plates were then exposed overnight with X-ray film (Kodak) and then developed.

Direction I	**Components**	**Ratio**	**Number of runs**
System			
A	Pet ether/ethyl acetate	98 :2	3
B	Pet ether/acetone	92:8	3
C	Chloroform/methanol	96:4	1
D	Chloroform/methanol/water	100:14:0.8	1
E	Chloroform/methanol/water	60:30:6	1
Direction II			
System			
A	Pet ether/acetone	98:2	1
B	Toluene/acetone	95:5	1
C	Toluene/acetone	80:20	1
D	Chloroform/acetone/methanol/water	50:60:2.3:3	1
E	Chloroform/aceticacid (glacial)/methanol/ water	40:25:3:6	1

Table 2.4. Solvents used to separate the different lipid components of the *M. tuberculosis* cell wall.

2.4.7 Transmission electron microscopy

M. tuberculosis rolling cultures were grown in Dubos to an OD_{600} of 0.6. A 20 ml aliquot of each strain to be used for microscopy was transferred to a 50 ml Falcon tube. Cells were centrifuged at 1,200 rcf for 15 minutes and thoroughly resuspended in 2 % glutaraldehyde / 2 % paraformaldehyde in 0.1 M sodium cacodylate buffer (pH 7.2) for 18 hours. The cells were prepared for transmission electron microscopy by Liz Hirst (NIMR). Cells were post fixed in 1 % osmium tetroxide then washed and stained en bloc with 1 % aqueous uranyl acetate for 2 hours then dehydrated. They were embedded in medium hardness Agar resin for 24 hours before polymerization at 70 ^{o}C overnight. 50 µm sections were mounted on pioloform coated slot grids and stained with saturated ethanolic uranyl acetate followed by Reynold's lead citrate. They were viewed in a Jeol 1200 EX transmission electron microscope and images were taken using an Orius TEM Charge Coupled Device (CCD) camera at x 25,000 magnification (Gatan Inc.).

2.4.8 ß-galactosidase transcription reporter assay

The ß-galactosidase reporter assay was used to assess the promoter activity of *pknF* under different conditions. The *pknF* promoter region was fused upstream of *lacZ* in pEJ414 (the plasmid is integrated into the genome of *M. tuberculosis* using pBSInt so only one copy is present per cell). The *lacZ* gene product is ß-galactosidase. In the presence of ß-galactosidase the substrate ortho-nitrophenyl-β-galactoside (ONPG) is hydrolysed to galactose and ortho-nitrophenol; the latter is yellow in aqueous solution and can be measured at 405 nm.

When required, a panel of drug and stress reagents were added to exponential phase cultures (OD_{600} of 0.6 to 0.8) for two hours at 37 ^{o}C with rolling (table 2.5). For acid stress, cultures at an OD_{600} of between 0.25 and 0.3 were harvested and then resuspended in either Dubos at pH 7.2 or Dubos adjusted to pH 5.5 and further incubated for 24 hours. All manipulations were then performed at 4 ^{o}C. 25 ml of exponential

(OD_{600} of 0.6 to 0.8) or stationary phase cultures (OD_{600} of 2) were harvested by centrifugation and resuspended in 10 ml cold Z buffer (no ß-mercaptoethanol). They were washed twice with 1 ml Z buffer (no ß-mercaptoethanol). Cells were then broken by ribolysis and filtered as described in section 2.4.4.

Next, to normalise the assay the amount of protein present in each sample was determined using the copper based bicinchoninic acid (BCA) protein assay kit (Pierce). The assay works in two steps. In the first step peptides chelate with copper to form a blue coloured complex from the reduction of cupric ions (Cu^{2+}) to cuprous ions (Cu^{1+}) under alkaline conditions. In the second step BCA chelates with the cuprous ion (Cu^{1+}) resulting in a purple colour which can be measured at 550 nm. 10 µl of each sample (in triplicate) was added to a 96 well microtitre plate. To the remainder of each sample 1.5 µl ß-mercaptoethanol (Sigma) was added and stored at 4 ^{o}C until required. Addition of β-mercaptoethanol stabilizes the ß-galactosidase enzyme. Fifty parts of BCA reagent A containing Bicinchoninic acid and tartrate in an alkaline carbonate buffer and 1 part BCA reagent B containing 4% copper sulfate pentahydrate solution were mixed, 200 µl of this mixture was added to each well, plates were incubated at 37 ^{o}C for 30 minutes and then absorbance was measured at 550 nm. A standard curve was generated using titrations of bovine serum albumin (BSA) of known concentrations.

The ß-galactosidase reporter assay was performed based on the method by Miller (1972). Assays were performed in biological triplicate. 100 µl of lysate was combined with 400 µl Z buffer plus 50 mM β-mercaptoethanol and equilibrated to 28 ^{o}C for five minutes. The reaction was started by addition of 100 µl of 15 mM ONPG. Colour formation was monitored, the reaction was stopped by addition of 250 µl 1M sodium carbonate when a yellow colour had developed and the time was noted. 300 µl of each sample (in duplicate) was added to a microtitre plate and absorbance was measured at 405 nm in a Benchmark plus microplate spectrophotometer (Bio-rad). The enzymes specific activity (units of β-galactosidase per mg of protein per minute) were then calculated using the following formula originally described by Miller (1972) : (380 x OD (405 nm))/(time (minutes) x volume of lysate used in assay (ml) x protein concentration (mg/ml)).

Drug/stress reagent	Final concentration
Izoniazid	0.0125 µg/ml
Ethambutol	0.375 µg/ml
Streptomycin	0.0625 µg/ml
Gentomycin	2 µg/ml
Mitomycin C	0.02 µg/ml
Ofloxacin	1 µg/ml
Hydrogen peroxide	2 mM
t-butyl hydrogen peroxide	0.1 mM
S-Nitrosoglutathione	5 mM
Diamide	10 mM
Plumbagin	0.5 mM
Sodium nitroprusside	10 mM

Table 2.5 The panel of drug and stress reagents used in this study to assess the promoter activity of *pknF* under different conditions.
The final concentration of each reagent used is equal to 50 percent of the minimum inhibitory concentration (MIC). Each reagent was added to the cultures for two hours.

2.4.9 *M. tuberculosis* growth inhibition (Alamar blue) assays

Growth inhibition assays were performed to assess the susceptibility of *M. tuberculosis* strains to various drugs and stress reagents based on the method by Collins and Franzblau (1997). The assay utilises the CellTiter-Blue reagent (Promega) which contains the compound resazurin. In the presence of metabolically active cells resazurin is reduced to resorufin, a fluorescent pink compound which can be measured with an excitation wavelength of 579 nm and emission of 584 nm.

Assays were carried out in 96 well plates (Greiner Bio-one). Two fold dilution series of the drug or stress reagent in 90 µl of Dubos were added to the plates. Positive (media plus bacteria) and negative (media only) control wells were included in all plates. Exponential phase *M. tuberculosis* cultures were diluted to give 1 x 10^4 cells in 10 µl. 10 µl of this dilution was added to the appropriate wells. Plates were incubated at 37 ^oC for five days to allow for *M. tuberculosis* growth. On day five 20 µl of CellTiter-Blue reagent was added to each well. Plates were incubated for 16 hours at 37 ^oC to allow cells to convert the resazurin to resorufin and then the fluorescent signal was measured in a Polarstar Galaxy plate reader (BMG) as per manufacturer's guidelines. Data were

plotted as the percentage inhibition of growth compared to the positive control (media plus bacteria) versus the reagent concentration.

2.4.10 Metabolic tests for *M. tuberculosis*: Phenotype Microarrays (BiOLOG)

Phenotype MicroArrays™ were used to assess cellular phenotypes (Biolog Inc., Hayward CA, USA). Each well of the 96 well plate arrays is designed to test a different phenotype. Phenotype microarray plates numbers one to ten containing metabolic tests for bacteria and fungi were used in this study. They included plates which tested the ability of *M. tuberculosis* to respire when exposed to a range of carbon, nitrogen, phosphorus and sulphur sources, various nutrient supplements, peptide nitrogen sources and osmolytes and also tested the ability of *M. tuberculosis* to respire in a range of pH conditions. Phenotype microarrays were initially set up at NIMR by Kathryn Lougheed.

20 ml of mid-log *M. tuberculosis* (OD_{600} 0.6-0.8) grown in Dubos was harvested by centrifugation. Cells were starved overnight by resuspension in 1 x PBS to ensure all traces of nutrients were metabolised to avoid false positive results in the assay. Cultures were harvested then resuspended in IF-0a base media (Biolog) recommended by the manufacturer and adjusted to an OD of 0.68. 0.88 ml was added to the Phenotype microarray stock solution which contained IF-0a base, phenotype microarray additive (which varied depending on which plate was used) and Dye Mix G containing a tetrazolium dye which is reduced in the presence of respiring cells forming a coloured compound. 100 µl of this mixture was added to each Biolog Plate well. Plates were incubated for seven days at 37 ^{o}C then absorbance was measured at 570 nm in a Polarstar Galaxy plate reader (BMG) as per the manufacturer's guidelines.

2.4.11 Analysis of small molecules and metabolites produced by *M. tuberculosis*

Culture supernatants of *M. tuberculosis* strains grown in Dubos or Sauton's medium were extracted at an OD_{600} of 0.3, 0.6, 1.0 and 2.0 by harvesting the cells and then filtering the supernatants twice through 0.22 µM filters as outlined in section 2.4.4. The metabolites present were then analysed by nuclear magnetic resonance (NMR) spectroscopy at Imperial College London by Volker Beherends (Dr Jake Bundy's

Laboratory). Briefly, 500 µl of sample plus 20 % deuterium, 5 mM sodium azide, 40 mM phosphate buffer and TSP was combined and added to a glass NMR vial then each sample was analysed in a Bruker DRX 600.

2.4.12 Murine bone marrow-derived macrophage infections

2.4.12.1 Differentiation of murine bone marrow derived macrophages (BMDMs)

All buffers and reagents were endotoxin free.

Bone marrow from six to eight week old female BALB/c immunocompetent mice was flushed into 10 ml of RPMI 1640 (Gibco, Invitrogen) supplemented with 10 % heat-inactivated foetal calf serum (Gibco, Invitrogen), 2 mM L-glutamine, 1 mM sodium pyruvate, 10 mM HEPES and 50 µM ß-mercaptoethanol (called RPMI complete) in a Petri dish. Cell clumps were broken using a syringe and then transferred to a 50 ml Falcon tube. Cells were centrifuged at 250 rcf for five minutes to pellet the cells. RPMI was removed and a 10 ml aliquot of 0.83 % ammonium chloride was added for 5 minutes at 37 ^{o}C to lyse the red blood cells. Cells were harvested again and the ammonium chloride was removed. Cells were washed in 10 ml 1 x PBS (Gibco, Invitrogen) then resuspended in 10 ml RPMI containing 20 % L cell conditioned medium (L-cell). L-cell is the supernatant produced from mouse fibroblast 929 cultures (prepared by the Large Scale Lab at NIMR) and contains colony stimulating factor which stimulates the differentiation of bone marrow derived monocytes into macrophages (Austin *et al.*, 1971). The cell solution was filtered through a 70 µm filter (BD Biosciences) to remove bone and muscle tissue. A 50 µl aliquot of the cell suspension was mixed with 50 µl trypan blue stain (Fluka) and 10 µl of this mix was counted using a Bright-Line haemocytometer (Sigma) counting chamber and an Axiovert 25 microscope (Zeiss). Monocytes were plated out in Petri dishes at a concentration of 4 x 10^6 cells per plate in RPMI plus 20 % L-cell supplement in a total volume of 10 ml. Cells were incubated at 37 ^{o}C with 5 % CO_2. After 72 hours a further 10 ml of RPMI plus 20 % L cell supplement was added to each Petri dish. After a further 48 hours macrophages were ready to use. All media and non-adherent cells were removed from the Petri dishes. 5 ml of cold (4 ^{o}C) 4 mM EDTA in 1 x PBS was added

to each Petri dish and a 10 ml stripette was used to gently flush the cells from the plate. Addition of EDTA helps to detach adherent cells as it is a chelating agent that binds calcium. Cells were transferred to a 50 ml Falcon tube and centrifuged at 250 rcf for five minutes. The supernatant was discarded and the cell pellet resuspended in RPMI containing 5 % L-cell supplement. A 50 µl aliquot of the cell suspension was mixed with 50 µl Trypan Blue and 10 µl of this mix was counted using a haemocytometer as before. 2 x 10^5 macrophages were seeded per well in a 1 ml volume in a 24 well cell culture plate (Tissue Culture Treated plates, Costar). When required, mouse interferon-γ (IFNγ) was added at 20 ng/ml (Roche) to activate the macrophages. Cells were plated out in triplicate for each *M. tuberculosis* strain and post-infection time point. Prior to infection cells were incubated for 18 hours at 37 ^oC with 5 % CO_2 to allow adherence and activation.

2.4.12.2 Preparation of bacterial strains and infection of BMDMs

M. tuberculosis rolling cultures were grown in Dubos medium to an OD_{600} of 0.6. A 20 ml aliquot of each strain to be used for infection was transferred to a 50 ml Falcon tube. Cells were centrifuged at 1,200 rcf for 15 minutes and thoroughly resuspended in PBS plus 0.05 % Tween 80. Cells were centrifuged at 50 rcf for five minutes to spin down any clumped cells and the supernatant was transferred to a new tube. The OD_{600} of each culture was determined and converted to CFU/ml (based on an OD_{600} of 1.000 being equivalent to 5 x 10^8 CFU/ml). *M. tuberculosis* strains were added to the macrophages in a 50 µl volume at a multiplicity of infection of one bacterium to two macrophages. Plates were incubated for a further six hours to allow the macrophages to phagocytose the bacteria. Extracellular bacteria were then removed from all the wells using an aspirator and cells were washed three times with 500 µl HBSS salt solution (Gibco). 1 ml RPMI complete containing 5 % L-cell was added to each well and cells were placed in a 37 ^oC incubator with 5 % CO_2. Survival and multiplication of *M. tuberculosis* within the macrophages was assessed at six, 24, 72, 120 and 168 hours post infection. At each time point media was removed from each well and macrophages were lysed in 500 µl dH_2O plus 0.05 % Tween 80 for 30 minutes. Appropriate dilutions of bacteria were then

plated onto 7H11 agar plates and incubated at 37 ^{o}C for two to four weeks. Colonies were then counted and CFU/ml was calculated.

2.4.12.3 Greiss nitrite assay

Macrophages are activated by IFN-γ secreted by T helper cells. Activated macrophages produce nitric oxide and oxygen radicals which both have antimicrobial activity. Nitrite (NO_2^-) is a stable breakdown product of nitric oxide and thus can be used as an indicator of macrophage activation using the Greiss assay which is based on a diazotization reaction. Under acidic conditions sulfanilamide and *N*-1-napthylethylenediamine dihydrochloride (NED) react with NO_2^- to give a coloured azo derivative which can be measured at 540 nm.

Culture supernatants from macrophage infections at various time points were filtered twice through 0.22 μM filters as in section 2.4.4. Assays were performed in 96 well flat bottomed plates (Greiner Bio-one). A standard curve was generated by diluting 0.1 M sodium nitrite ($NaNO_2$) 1:1,000 in supplemented RPMI to give a 100 μM solution of NO_2^-. A two fold dilution series was performed to produce the standard curve. 50 μl of each experimental sample was added to the wells (in triplicate). 50 μl of Solution A (1 % sulfanilamide, 5 % phosphoric acid) was then added to both experimental samples and standards and plates were incubated at room temperature in the dark for 10 minutes. 50 μl of Solution B (0.1 % NED) was then added to all wells and incubated as before. Absorbance was then measured at 540 nm in a Benchmark plus microplate spectrophotometer (Bio-rad).

2.4.13 Preparation of bacterial strains for infection in a mouse aerosol model

M. tuberculosis rolling cultures were grown in Dubos to an OD_{600} of 0.6. Cultures were pelleted and resuspended in DMEM (Sigma) plus 2 mM L-Glutamine and 50 % foetal calf serum to a concentration of 10^9 bacteria per ml. Aerosol infections were performed at the National Institute for Biological Standards and Control by Dr. Barry Walker and Dr. Angela Rodgers.

2.4.14 Enzyme-linked immunosorbent assays (ELISAs)

Two ELISA methods were employed; a cytokine ELISA and *M. tuberculosis* whole cell ELISA. Both methods are outlined below. They each used Maxisorp F96 Nunc-Immuno plates (Nunc) whose surface has high affinity for molecules with mixed hydrophilic and hydrophobic domains.

1. Cytokine ELISAs

Cytokine ELISAs were used for the measurement and quantification of cytokines secreted by macrophages post infection with *M. tuberculosis*. Murine bone marrow derived macrophages were differentiated and *M. tuberculosis* strains were prepared as in section 2.4.4.12. Post infection, macrophages were not washed and media was not replaced. Instead, at six, 24, 72 and 96 hours post infection the media was filtered twice through 0.22 μM membranes as in section 2.4.4 to remove extracellular *M. tuberculosis* and then ELISAs were performed. All incubation steps were performed at room temperature unless stated otherwise. Briefly, 50 μl of desired capture antibody in Coating Buffer (eBioscience) was immobilized onto Maxisorp plates overnight at 4 ^{o}C. Plates were washed three times in 1 x PBS plus 0.05 % Tween 20 (Sigma). Wells were then blocked by incubation for one hour with 200 μl/well of 1 x Assay Diluent (eBioscience). Wells were then washed as previously. Standards were diluted in 1 x Assay Diluent to make the standard curve and added to each plate. 50 μl of each filtered sample were then added to the plate (in triplicate) and incubated for two hours. Wells were washed as previously then 50 μl/well of biotin-conjugated detection antibody (eBioscience) diluted in 1 x Assay Diluent were added and incubated for one hour. Wells were washed and then 50 μl/well of Avidin-HRP diluted in 1 x Assay Diluent were added and incubated for 30 minutes. Wells were then washed seven times as previously and developed by addition of 50 μl 3,3′,5,5′-Tetramethylbenzidine (TMB) substrate for horseradish peroxidase (Sigma). To stop the reaction 50 μl 1M H_2SO_4 was added and plates were measured at 450 nm in a Benchmark plus microplate spectrophotometer (Bio-Rad).

2. *M. tuberculosis* whole cell ELISAs

Mid-log rolling cultures of *M. tuberculosis* were harvested, resuspended in PBS and heat killed for one hour at 80 ^{o}C. 50 µl of the cell suspension was applied to each well and incubated at room temperature for two hours to allow adsorption to the bottom of each well. Wells were then blocked overnight at 4 ^{o}C with 200 µl of 1 % BSA in TBS (50 mM Tris-HCl pH8, 150 mM NaCl) to reduce non-specific binding of primary antibodies. Primary antibodies against LAM and various capsular components of the *M. tuberculosis* cell wall were raised in mice (IgM purified) and were kindly provided by Drs Ben Appelmelk and Jeroen Geurtsen (VU University Medical Centre, Amsterdam). Mab F183-24 primarily recognises PIM6, but also ManLAM capped with 3 mannosyl residues; Mab F30-5 recognises the Ara6 structure present in LAM; Mab Baba recognises capsular glucan and 55.92.1a1 recognises ManLAMs capped with single mannosyl residues. Plates were washed three times in 1 x PBS plus 0.05 % Tween 20 (Sigma). Appropriate dilutions of primary antibody in 50 µl volumes in TBS were added to each well and incubated at 37 ^{o}C for 90 minutes. Wells were washed as before then 50 µl of a 1:10, 000 dilution of secondary antibody (Goat Anti-Mouse IgM (µ-chain specific) peroxidase conjugate (Sigma)) was added to each well and incubated as previously. Wells were washed and developed by addition of 50 µl 3,3´,5,5´-Tetramethylbenzidine (TMB) substrate for horseradish peroxidase (Sigma). Colour formation was monitored (wells turn blue in the presence of secondary antibody) and the reaction was stopped by adding 50 µl 1M H_2SO_4 to each well producing a yellow coloured product. Plates were then measured at 450 nm in a Benchmark plus microplate spectrophotometer (Bio-Rad).

2.5 Protein Expression and Analysis Techniques

2.5.1 Sodium dodoecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

All SDS-PAGE experiments were performed with eight to 15 % Acrylamide/Bis-Tris gels cast in 1 mm Casettes (Invitrogen) according to Sambrook (Sambrook & Russell, 2001). Protein samples were mixed with 4 x Laemmli Sample Loading Buffer (0.25 M Tris pH 6.8, 6 % SDS, 40 % glycerol, 0.04 % Bromophenol blue, 20 % ß-

mercaptoethanol) and heated to 100 ^{o}C for five minutes prior to loading. Five to 20 µl of each sample was typically loaded alongside a Dual Color Precision Plus Protein Standard (Bio-Rad). Gels were run in 1 x SDS Running Buffer (14.4 g Glycine, 3 g Tris, 1 g SDS) at 170 to 200 V for 45 to 60 minutes in an XCell SureLock mini-cell (Invitrogen). Gels were developed by staining with InstantBlue Coomassie stain (Expedeon) according to the manufacturer's instructions and imaged on a BioDoc-It Imaging System.

2.5.2 Expression and purification of *M. tuberculosis* proteins in *E. coli*

2.5.2.1 Protein expression

The pGex-6P-1 vector (GE Healthcare) was chosen for all protein expression studies. pGex-6P-1 has an N-terminal Glutathione S-Transferase (GST) tag which binds to glutathione sepharose allowing for affinity purification of target proteins. A site cleavable with PreScission Protease (GE Healthcare) is situated 3′ to the GST tag to allow for removal of the GST. Protein expression is under control of the *tac* inducible promoter (situated upstream of the GST gene) which is a hybrid of two strong promoters, *trp* and *lac*. The *tac* promoter is induced with isopropyl beta-D-thiogalactoside (IPTG). In addition, pGex-6P-1 also contains an internal *lac*I gene. LacI is a repressor protein which binds to the *tac* promoter in the absence of inducer thus allowing for tighter control of protein expression.

All expression studies used *E. coli* BL21 (DE3) star competent cells (Invitrogen) or *E. coli* BL21 (DE3) pRep4 (V. Molle, Personal Communication) cells which contain an extra chaperone expressing plasmid to enhance protein stability. One colony from an overnight transformation plate was picked into L. broth containing appropriate selection antibiotics and grown with shaking (250 rpm) overnight at 37 ^{o}C. 0.01 volumes of overnight culture was inoculated into fresh L. broth (usually two to four litres) and grown as previously until the OD_{600} reached 0.6 (mid-log phase). Protein expression was then induced with 0.1 mM IPTG for 16 hours at 4 ^{o}C. All subsequent steps were performed at 4 ^{o}C unless stated.

Cells were harvested by centrifugation for 20 minutes at 13,500 rcf, washed in PBS then resuspended in 10 ml of lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, plus one Complete EDTA-Free Protease Inhibitor Cocktail Tablet (Roche) per 50 ml) per gram of cells. When required, 10 % glycerol was added to all protein buffers to increase protein stability. DNase, RNase (Promega) and lysozyme (Sigma) were added to the cells to aid lysis and reduce viscosity. Cells were lysed by sonication (Vibracell, Sonics) on ice with ten bursts of ~ 20 seconds at amplitude 10. Cell lysates were then centrifuged for one hour at 100,000 rcf to separate the soluble from the insoluble fraction.

2.5.2.2 Protein Purification

The soluble lysate was applied to an appropriate amount of prepared Glutathione Sepharose 4B resin (three to five ml) (GE Healthcare) equilibrated in lysis buffer. It was incubated overnight at 4 ^{o}C with gentle mixing. The resin was collected by centrifugation at 400 rcf for five minutes then washed with one litre of wash buffer (50 mM Tris (pH 8), 500 mM NaCl). At this stage two different methods were employed. If an un-cleaved GST-fusion protein was required the target protein was eluted at this stage with wash buffer supplemented with 10 mM reduced glutathione (Sigma) by incubation overnight with mixing as previously. Alternatively, if a cleaved protein was required washed resin was equilibrated with PreScission protease cleavage buffer (50 mM Tris (pH 8), 200 mM NaCl, 1mM EDTA, 0.5 mM DTT). 50 µl of PreScission protease was then added to the resin and incubated overnight with mixing as previously. Elutions were concentrated in a 20 ml Vivaspin ultrafiltration concentrator (VivaScience) with an appropriate molecular weight cut off filter (three to 10 kDa). Samples were concentrated by centrifugation at 2,000 rcf.

2.5.2.2.1 Size exclusion chromatography

To remove any protein contaminants in the preparations each sample was purified by size exclusion chromatography with a pre-packed HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) using an AKTA Prime system. The column was equilibrated in gel filtration buffer (40 mM Tris (pH 8), 200 mM NaCl), the sample was loaded and then run at between 0.5 and 1 ml per minute collecting fractions. Appropriate

fractions containing the protein of interest were checked by SDS-PAGE, pooled, concentrated and snap-frozen at -80 ^{o}C in 20 to 100 µl aliquots.

2.5.2.3 Mass spectrometry

Proteins were identified by MALDI-TOF and electrospray mass spectrometry. For MALDI-TOF the protein band was gel excised, cut into 2 mm x 2 mm pieces, then 500 µl 200 mM ammonium bicarbonate (ABC) and 50 % acetonitrile was added for 30 minutes at 25 ^{o}C to extract SDS and Coomassie Blue. The protein was then reduced in 200 µl 20 mM DTT, 200 mM ABC and 50 % acetonitrile for one hour at 25 ^{o}C. The gel pieces were then washed extensively in 500 µl 200 mM ABC and 50 % acetonitrile buffer. Cysteines were alkylated in 5 mM iodoacetamide, 200 mM ABC and 50 % acetonitrile for 1 hour at 25 ^{o}C in the dark and then washed extensively with 20 mM ABC in 50 % acetonitrile. 500 µl neat acetonitrile was then added for 10 minutes. All liquid was then removed and the gel pieces allowed to dry. They were then trypsin digested for 24 hours at 37 ^{o}C by dissolving trypsin (Promega) in 5 mM ABC to a final concentration of 2 µg/ml and adding 35 µl to the gel pieces. MALDI was then performed by Steven Howell (NIMR). For electrospray ionisation (ESI) mass spectrometry 100 pmols of purified protein were analysed by Steven Howell (NIMR).

2.5.3 Biochemical techniques

2.5.3.1 *In vitro* phosphorylation assays

In vitro phosphorylation was carried out in 20 µl reactions containing the recombinant kinase (1 µg), substrate (1 µg) (if necessary) and 200 µCi/ml [γ -32P] ATP (Perkin Elmer) in phosphorylation buffer (25 mM Tris-HCl pH 6.8, 1 mM DTT, 5 mM $MgCl_2$, 1 mM EDTA). The reaction was carried out for 30 minutes at 37 ^{o}C. The reaction was stopped by addition of Laemmli Sample Loading Buffer and incubated at 100 ^{o}C for five minutes before analysis by SDS-PAGE. After electrophoresis gels were washed in 10 % trichloracetic acid for 10 minutes at 90 ^{o}C then stained with InstantBlue Coomassie stain (Expedeon) or dried on a Gel Drier (Savant, Slab Drier) then exposed to a

phosphoimager screen (Molecular Dynamics) and developed on a Storm Scanner (Amersham).

2.5.3.2 Liquid chromatography / mass spectrometry / mass spectrometry (LC/MS/MS)

LC/MS/MS was performed in collaboration with V. Molle (IBCP, France) to identify residues in Rv1747 that were phosphorylated by PknF. Recombinant $PknF^{1-292}$ and $Rv1747^{1-559}$ was supplied to V. Molle for the phosphorylation site identification. LC-MS/MS was performed as outlined in two papers (Canova *et al.*, 2008; Molle *et al.*, 2006). Briefly this technique involved the *in vitro* phosphorylation of Rv1747 by PknF with ATP and then the phosphorylated Rv1747 protein was separated from the kinase by liquid chromatography. Phosphorylated Rv1747 peptides were identified by on-target dephosphorylation/MALDI-TOF which is where a phosphatase is used to remove phosphoryl groups from peptides leading to de-phosphorylated peptides appearing as new or more abundant peaks on the MS spectra (Molle *et al.*, 2006). LC/MS/MS was then performed to reveal the location of the phosphorylation sites. This technique uses liquid chromatography (LC) to separate the pool of phosphopeptides which helps in the identification of phosphorylation sites by reducing the complexity of the mixture, coupled with MS for the identification of the phosphorylated residues (Molle *et al.*, 2006). Repeated rounds of mass spectrometry were required to identify multiple phosphorylation sites.

2.5.3.3 Isothermal titration calorimetry (ITC)

ITC was used to determine the binding kinetics between FHA domains and potential phosphopeptide substrates. Experiments were performed with the MicroCal iTC_{200} system (GE Healthcare). The instrument can determine the binding affinity (K_A), enthalpy change (ΔH) and binding stochiometry (n) between two reactants in a single experiment. The aim of ITC is to calculate the dissociation constant (K_D) which measures how tightly a ligand binds to a particular protein. Lasse Stach (NIMR) provided expert help with the use of the instrument and phosphopeptides were synthesised by Dr. W. Mawby (University of Bristol).

The instrument is composed of a reference cell and sample cell. In the reference cell is an aqueous solution. Both reference cell and sample cell were cooled to 10 ^{o}C while the surrounding buffer was cooled to 9 ^{o}C so that the instrument needed to use a specific amount of energy to keep the reference cell solution at 10 ^{o}C. 200 µl of FHA protein (at 50 µM) in 20 mM Tris pH8 and 150 mM NaCl was loaded into the sample cell. 100 µl of phosphopeptide (at 500 µM) in the same buffer was loaded into the syringe. During the experiment 2 µl of ligand was titrated into the sample cell every 180 seconds with a total of 20 injections. Binding of the protein to the peptide can be an exothermic or endothermic reaction (due to the net outcome of hydrogen bond formation and electrostatic interactions). The raw data for an experiment consisted of a series of spikes which were plotted as the amount of power (in µcal/sec) needed to maintain the reference and sample cell at the same temperature. Each spike represented one injection of ligand. The total heat effect per injection was then calculated by integrating each spike with respect to time. From this, the dissociation constant was determined. The K_D is defined as the concentration of ligand at which the binding site on a particular protein is half occupied, hence the smaller the dissociation constant, the more tightly bound the ligand.

2.5.3.4 Size-Exclusion Chromatography Combined with Multiangle Laser Light Scattering (SEC-MALLS).

SEC-MALLS is a technique which combines size exclusion chromatography and light scattering to measure the absolute molecular mass and concentration of proteins of interest. Briefly, protein samples are separated by size using a gel filtration column before passing through light scattering, absorbance and refractive index detectors. When a laser light hits a protein it causes light to scatter, the intensity of which is directly proportional to the molar mass and can be shown by the following equation:

Intensity of scattered light = molecular weight x concentration of protein (mg/ml)

Light scattering was measured at 90^{o}. For proteins over 500 kDa or with unusual structures multi-angle laser light scattering is necessary to accurately calculate the mass as scattering varies greatly with each angle measurement but this was not necessary in

this study. The concentration of the protein sample can be measured using absorbance or differential refractive index detectors. For this study the differential refractive index was calculated as the protein construct does not absorb UV due to an absence of tyrosine, tryptophan and phenylalanine residues. The refractive index detector is formed of two cells of which one acts as a reference cell (Optilab rEx). Both cells are conditioned with buffer. When a protein sample passes through the sample cell the difference in the refractive index between the two cells is measured and this has been shown to be directly proportional to the concentration of protein. Therefore measuring the molecular mass of each protein peak is possible by combining measurements from light scattering and the differential refractive index.

SEC-MALLS was used to investigate whether $FHA2^{202-310}$ became dimeric when phosphorylated at the N-terminus of the construct (at residue T208) by PknF. $FHA2^{202-310}$, $PknF^{1-292}$ and GST-$PknF^{1-292}$ were expressed and purified to homogeneity as described in section 2.5.2. Phosphorylation reactions (250 µl volume) were performed for four hours at 37 ^{o}C containing $FHA2^{202-310}$ and GST-$PknF^{1-292}$ (at a 10:1 ratio) with 500 µM ATP in phosphorylation buffer (25 mM Tris-HCl pH 6.8, 1 mM DTT, 5 mM $MgCl_2$, 1mM EDTA). $FHA2^{202-310}$ was typically used in the range of 1 to 5 mg/ml. Negative controls containing no ATP were also included. After the four hour incubation prepared Glutathione Sepharose 4B resin (100 µl) (GE Healthcare) equilibrated in phosphorylation buffer was added to each sample and mixed on a rotator for 1 hour at 4 ^{o}C to allow GST-$PknF^{1-292}$ to bind to the resin. Samples were then filtered through low binding 0.22 µm Durapore membrane filters (Ultrafree-MC, Millipore) to remove resin containing the bound GST-$PknF^{1-292}$. Next, 100 µl volumes of each sample were loaded onto a Superdex 200, 10/300 GL column (at 25 ^{o}C) conditioned with phosphorylation buffer plus NaCl (25 mM Tris-HCl pH 6.8, 150 mM NaCl, 1 mM DTT, 5 mM $MgCl_2$, 1mM EDTA, 3 mM sodium azide) at a speed of 0.5 ml/minute before passing through UV, refractive and light scattering detectors (Dawn Heleos and Optilab rEx, Wyatt). Data were analysed and graphs were drawn using JASCO ChromPass Chromatography data system and ASTRA [EASIGraph] plotting the retention time (in minutes) of each protein versus the molar mass (g/mol). Lasse Stach (NIMR) helped in the analysis of

data. Electrospray mass spectrometry was also carried out to confirm the phosphorylation of $FHA2^{202-310}$ as described in section 2.5.2.3.

CHAPTER 3. CHARACTERISATION OF THE *pknF-Rv1747* OPERON, INVESTIGATION INTO THE *pknF* STIMULUS AND INITIAL ASSESSMENT OF THE *pknF* AND *Rv1747* MUTANT STRAINS

3.1 Introduction

This study initially set out to define the operon structure of the genomic region from *pknF* to *Rv1749c* and to map the transcriptional start site of *pknF*. *pknF* lies within a putative operon with *Rv1747*. Co-transcription of genes can suggest that proteins act together to serve a particular function. In the case of the FHA domain containing proteins of *M. tuberculosis*, four of the six FHA proteins, including Rv1747, lie in putative operons with serine-threonine protein kinases. They include *embR* (*Rv1267c*) which forms a putative operon with *pknH*. Studies have demonstrated that EmbR, a transcriptional regulator, is phosphorylated by PknH (Molle *et al.*, 2003). Furthermore, phosphorylation by PknH increases the DNA binding capacity of EmbR to the promoter regions of the *embCAB* genes thus increasing their transcription (Sharma *et al.*, 2006). The *embCAB* genes are arabinosyltransferases which are involved in the arabinosylation of lipomannan to make lipoarabinomannan and the arabinosylation of arabinogalactan, two key virulence associated components of the mycobacterial cell wall (Sharma *et al.*, 2006). In addition, the FHA domain containing proteins *fhaA* (*Rv0020c*) and *fhaB* (*Rv0019c*) both lie in the same putative operon with *pknA* (*Rv0015c*) and *pknB* (*Rv0014c*). *pknA* and *pknB* are essential genes in *M. tuberculosis* and have been shown to play a role in cell wall synthesis and cell morphology (Kang *et al.*, 2005). PknB is known to phosphorylate both Rv0019c and Rv0020c *in vitro* (Grundner *et al.*, 2005; Gupta *et al.*, 2009). Therefore it was of interest to assess whether *pknF* and *Rv1747* were part of an operon.

Transcriptional start site mapping can yield insights into how gene expression is regulated by indicating the position and number of promoters present for genes of interest. Mapping transcriptional start sites also aids in the design of promoter reporter plasmid constructs for experiments investigating the stimulus of a particular gene.

Therefore, when the transcriptional start site of *pknF* was mapped by 5´RACE, experiments investigating the possible stimulus of *pknF* at the transcriptional level were performed using a promoter-*lacZ* fusion construct containing the *pknF* promoter region. It was hypothesised that upon subjecting the *M. tuberculosis* cell to a particular stress the *pknF* promoter activity may increase, presumably through the binding of a transcription or sigma factor to the promoter, leading to a greater abundance of *pknF* and *Rv1747* transcript, which may in turn be translated into an increased amount of protein. Autophosphorylation of PknF might then occur after sensing the stress signal making the protein functionally active, and then sequentially PknF may phosphorylate Rv1747 in order to transduce the signal and regulate the activity of the transporter. In 2008 Jayakumar and colleagues demonstrated that the *pknE* promoter showed increased activity when *M. tuberculosis* cells were exposed to nitric oxide stress (Jayakumar *et al.*, 2008). They went on to demonstrate that a *pknE* mutant conferred increased resistance to nitric oxide donors and related it to a mechanism where *pknE* senses nitric oxide stress and prevents apoptosis in the host cell by interfering with signalling mechanisms. Therefore it was of interest to investigate whether the *pknF* promoter showed any change in activity when *M. tuberculosis* cells were exposed to a variety of stresses.

Initial characterisation of the *pknF* and *Rv1747* null mutants showed their growth was the same as the wild type strain under *in vitro* growth conditions (R. Whalan, Personal communication and Curry *et al.*, 2005). The next aim in this study was to further characterise both mutants by assessing their susceptibility to a range of drug and stress reagents compared with the parental wild type H37Rv strain. This was achieved by performing a series of 96 well plate growth inhibition assays. It was hypothesised that any differences in susceptibility between the strains may yield clues as to the function of the two proteins.

3.1 Experiments to determine co-transcription of *pknF* and *Rv1747*

The genomic localization of *pknF* and *Rv1747* (figure 3.1a) suggested that the two genes could potentially form an operon; the intergenic region between *pknF* and *Rv1747* is only 62 base pairs in size. RT-PCR experiments were performed as outlined in section

2.3.1 to answer this question. The results, shown in figure 3.1b, clearly show a transcript present running from *pknF* through the intergenic region into *Rv1747* demonstrating that these two genes are indeed co-transcribed. Primers used in the RT-PCR study can be found in Appendix II. A transcript was also present for each investigated gene when using gene specific internal primers. No transcript was detected in the intergenic region between *Rv1747* and *Rv1748* when using a forward primer within *Rv1747* with a reverse primer within *Rv1748*, indicating that these two genes are not co-transcribed. In addition a transcript was not present between *Rv1748* and *Rv1749c* which agrees with theory as the two genes are convergent. **The results conclusively showed that *pknF* and *Rv1747* form a single operon**.

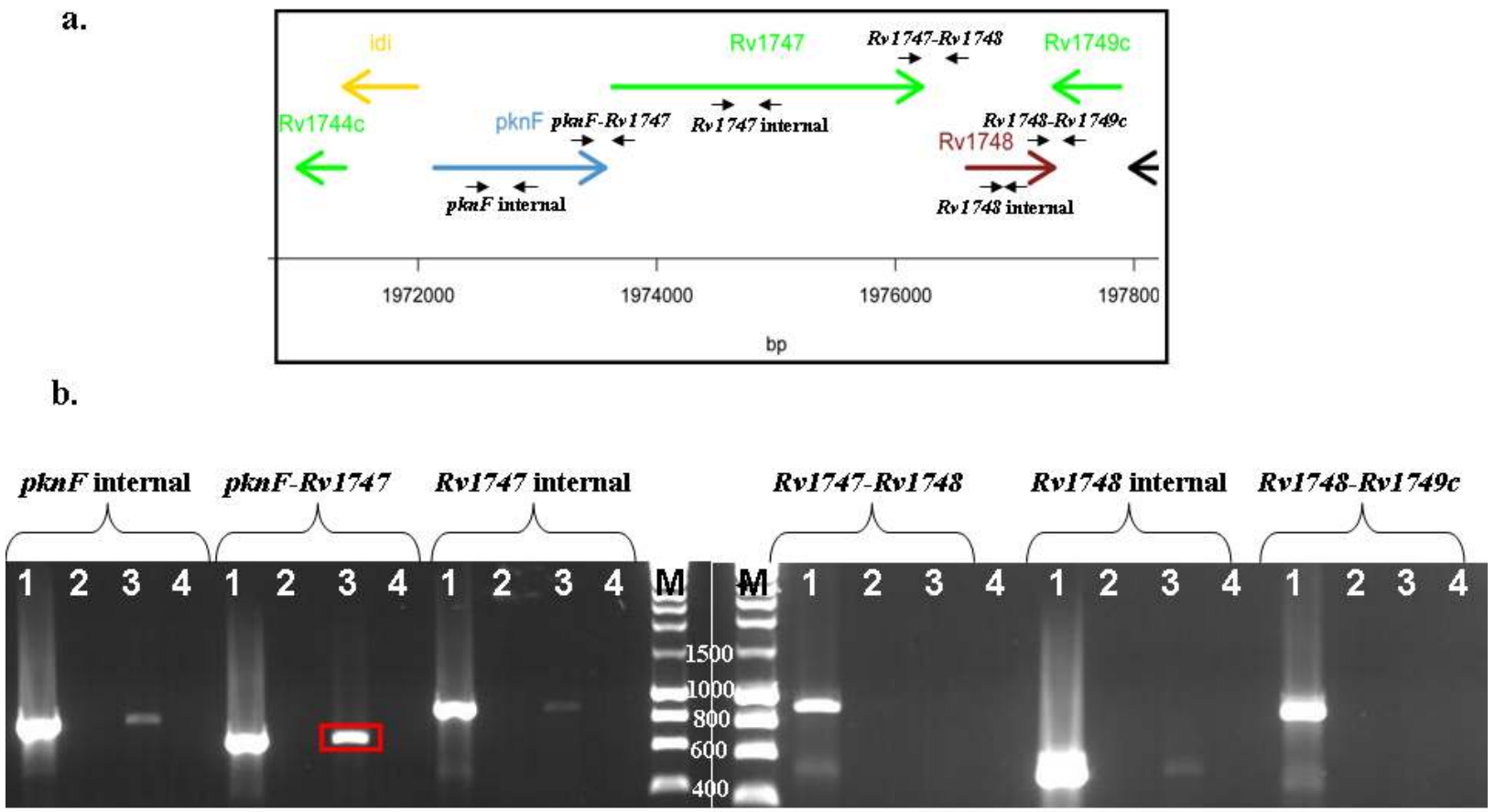


Figure 3.1 *pknF* and *Rv1747* are co-transcribed. **a**. Schematic showing the positions of *pknF*, *Rv1747* and the surrounding genomic region of *M. tuberculosis* (adapted from Tuberculist (http://tuberculist.epfl.ch)). The marks on the chromosome are at 2000 bp intervals **b**. RT-PCR results showing the products of cDNA amplification. 10 µl of a 50 µl RT reaction was analysed by agarose gel electrophoresis. Labels above the gel show the primer pairs used in each PCR (Appendix II). Lane one of each primer pair is the positive control (H37Rv genomic DNA). Lane two is a negative control (no mRNA added to the RT reaction). Lane three is the RT positive lane. Lane four is RT negative (no reverse transcriptase present in RT reaction). The *pknF-Rv1747* transcript is highlighted with a red box. M is the DNA ladder and the relevant sizes in base pairs are labelled.

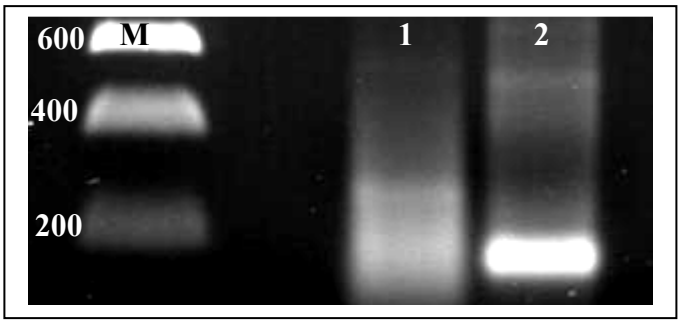
3.2 Determination of the transcriptional start site of *pknF*

5´RACE is a technique in which cDNA is generated from RNA and then PCR amplification is used to map the 5´ end of transcripts thus revealing the position of transcriptional start site/s (described in section 2.3.2).

The transcriptional start site of *pknF* was mapped; figure 3.2a shows the products of the primary and secondary PCR reactions. The product from the secondary PCR reaction was gel purified, cloned into pCR4Blunt-TOPO and six resulting transformants were sequenced; figure 3.2b summarises the 5´RACE results highlighting the possible transcriptional start sites. As 5´RACE incorporates a poly-cytosine tail onto the 3´ end of cDNA this technique cannot precisely map sites where transcription may begin at a cytosine residue. Therefore from the sequencing results it was not possible to determine whether transcription starts from either of the cytosine residues highlighted in blue or whether transcription is in fact coupled to translation and hence the start site is the adenine shown in red (figure 3.2b). The sequencing results demonstrate that **the transcriptional start site of *pknF* is most probably the same as the translational start site**.

For this study it was important to determine the transcriptional start site of *pknF* to aid in the design of the promoter-*lacZ* fusion construct which was to be used to investigate the possible stimulus to which the *pknF* promoter may respond. The *pknF* promoter would most likely be situated in close proximity upstream of the transcriptional start site. Therefore by using 5´RACE to narrow down the possible start site to a very small region it helped in designing the promoter-*lacZ* fusion construct.

a



b

5′ cctgccctcctggcgggctacgatgccg**cc**atgccgctcgcggaaggt 3′

Figure 3.2 The transcriptional start site of *pknF*. a. 5′ RACE results of *pknF* showing the products of the primary and secondary PCR reactions. 10 µl of each PCR product was analysed by agarose gel electrophoresis. M is the DNA ladder and the relevant sizes in base pairs are labelled. Lane 1 shows the products of the primary PCR amplification and lane 2 shows the products of the secondary nested PCR reaction which map the transcriptional start site. **b**. Summary of the 5′ RACE results mapping the possible transcriptional start site/s of *pknF* (highlighted in blue) and the translational start site according to the Institute Pasteur in Tuberculist (shown in red) (http://tuberculist.epfl.ch/).

3.3 Investigation of the transcription of the *pknF-Rv1747* operon after stress

PknF is a transmembrane protein kinase that contains an intracellular kinase domain and an extracellular domain thought to be responsible for signal recognition. The external stimulus that PknF recognises is unknown; it is possible that this signal influences *pknF* transcription and as a consequence was investigated at the transcriptional level in promoter-*lacZ* reporter gene assays utilising ß-galactosidase activity as a measure of promoter activity (section 2.4.8). One hypothesis was that upon subjecting the *M. tuberculosis* cell to a particular stress the *pknF* promoter activity may increase, leading to a greater abundance of *pknF* and *Rv1747* transcript, which may in turn be translated into a greater amount of functional protein. A promoter-*lacZ* fusion construct was generated by cloning 321 bp upstream of the *pknF* transcriptional start site in front of the *lacZ* gene in pEJ414 (there are 146 bp between *pknF* and its upstream divergent gene *idi*). This construct was sequenced and confirmed to be correct before electroporating into *M. tuberculosis*. Three separate *M. tuberculosis* colonies transformed with either *pknF*-pEJ414 or pEJ414 vector control were assessed for their promoter activity as described in section 2.4.8. Figure 3.3 shows that the *pknF* promoter in this assay had a specific activity of 320 units of ß-galactosidase whilst the pEJ414 vector control strain had a specific activity of 10 units. **These results confirm the presence of the *pknF* promoter within the construct and demonstrate that the promoter is active during exponential growth *in vitro***.

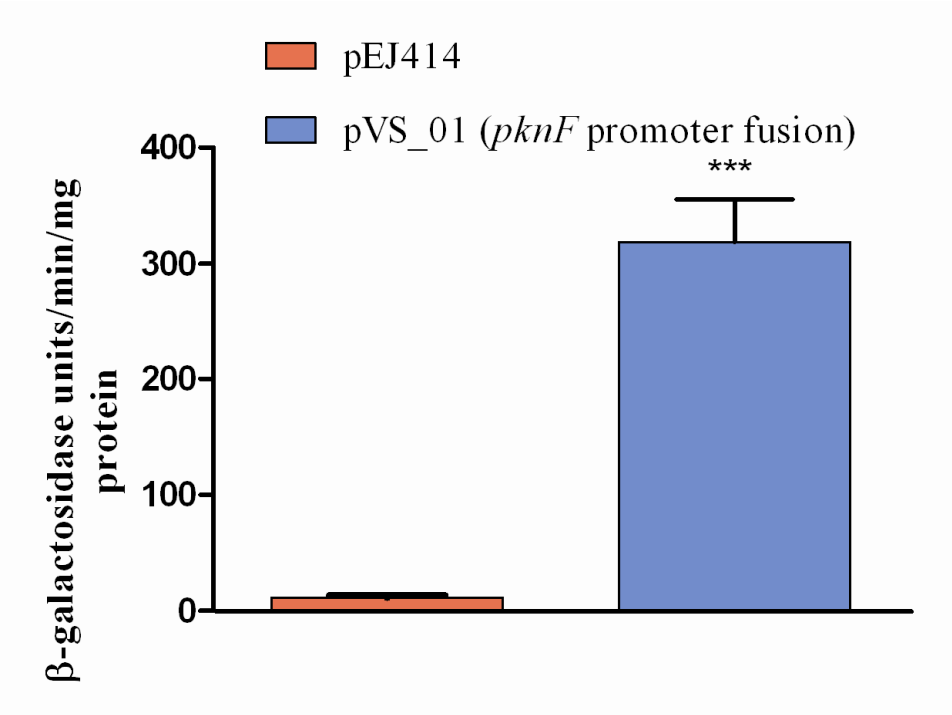


Figure 3.3 ß-galactosidase assay on the *pknF* promoter-*lacZ* strain and pEJ414 control strain in *M. tuberculosis*. The ß-galactosidase assay was performed on CFE of *M. tuberculosis* strains containing pEJ414 and pVS_01. The data presented are the means of three biological replicates each assayed in duplicate and the error bars represent standard deviations. An unpaired t-test (two-tailed) was used for a pairwise comparison between pEJ414 and pVS_01. The ß-galactosidase activity levels were determined to be significantly higher within strains containing pVS_01 compared with the control strain pEJ414. *** highlights the significant P value of <0.0001.

Next, to investigate whether transcription of the *pknF-Rv1747* operon was affected by external conditions ß-galactosidase assays were performed to assess promoter activity after treating cells for two or 24 hours with a variety of drugs and stress reagents as described in section 2.4.8.

The panel of stress reagents tested in this study included the first line anti-tuberculosis drugs isoniazid and ethambutol. The prodrug isoniazid is activated by the catalase peroxidise enzyme KatG and then binds to InhA resulting in inhibition of mycolic acid synthesis. Ethambutol acts on the biosynthesis of arabinogalactan, the major polysaccharide of the TB cell wall, by inhibiting the polymerization of arabinan. Both these drugs were chosen as it was hypothesised that Rv1747 could be involved in transporting a component of the cell wall and so the promoter of *pknF* may be induced if faced by cell wall stresses.

Gentamycin and streptomycin fall into the aminoglycosides category of antibiotics. They bind to the 30S ribosomal subunit blocking further protein synthesis. Hydrogen peroxide, t-butyl hydrogen peroxide and diamide are oxidative stress-inducing agents and treatment of cells with plumbagin results in superoxide stress. GSNO and sodium nitroprusside are nitric oxide donors and were chosen for investigation along with hydrogen peroxide based on the earlier *pknE* promoter study (Jayakumar *et al.*, 2008).

Mitomycin C is a DNA damaging agent that crosslinks DNA thus blocking replication. Ofloxacin is a fluoroquinolone which can also be considered to cause DNA damage as it interferes with DNA replication by inhibiting DNA gyrase function necessary for cell division. All these compounds were tested for their ability to alter the ß-galactosidase activity of the *pknF* promoter region in order to indicate a possible role for this kinase in *M. tuberculosis*. Isoniazid, ethambutol, gentomycin, streptomycin, hydrogen peroxide, t-butyl hydrogen peroxide, mitomycin C, ofloxacin, s-nitrosoglutathione and sodium nitroprusside were added to *M. tuberculosis* cells at an OD_{600} of 0.6 to 0.8 for two hours at concentrations equal to 50 percent of the published minimum inhibitory concentration (MIC) and CFEs were made as detailed in section 2.4.8. The ß-galactosidase activity of *pknF* after a 24 hour incubation in acidified medium to generate acid stress was also assessed as was the level of activity in stationary phase of growth (OD_{600} of 2.0).

The results from this study are shown in figure 3.4. The data labelled 'control, no stress' represent the results shown in figure 3.3 where CFE was generated from cultures in exponential phase of growth after having no form of treatment. All other data are labelled with the treatment that they were given. In all cases the ß-galactosidase activity levels were significantly higher within strains containing pVS_01 compared with the control strain pEJ414. **The results show that there are no significant changes in the levels of ß-galactosidase activity upon treatment with any of the reagents tested**. In addition, no significant changes were observed between the control strains after treatment. Data are displayed as the mean of three biological replicates. Data labelled 'pH7.2' represents the control experiment for the acid stress treatment labelled 'pH5.5'. As outlined in section 2.4.8 these cultures were harvested at an OD_{600} of 0.3 and then resuspended in either normal Dubos medium at a pH of 7.2 or Dubos medium adjusted to a pH of 5.5 to induce acid stress. Furthermore, the activity of the *pknF* promoter in stationary phase is comparable to that in exponential phase of growth.

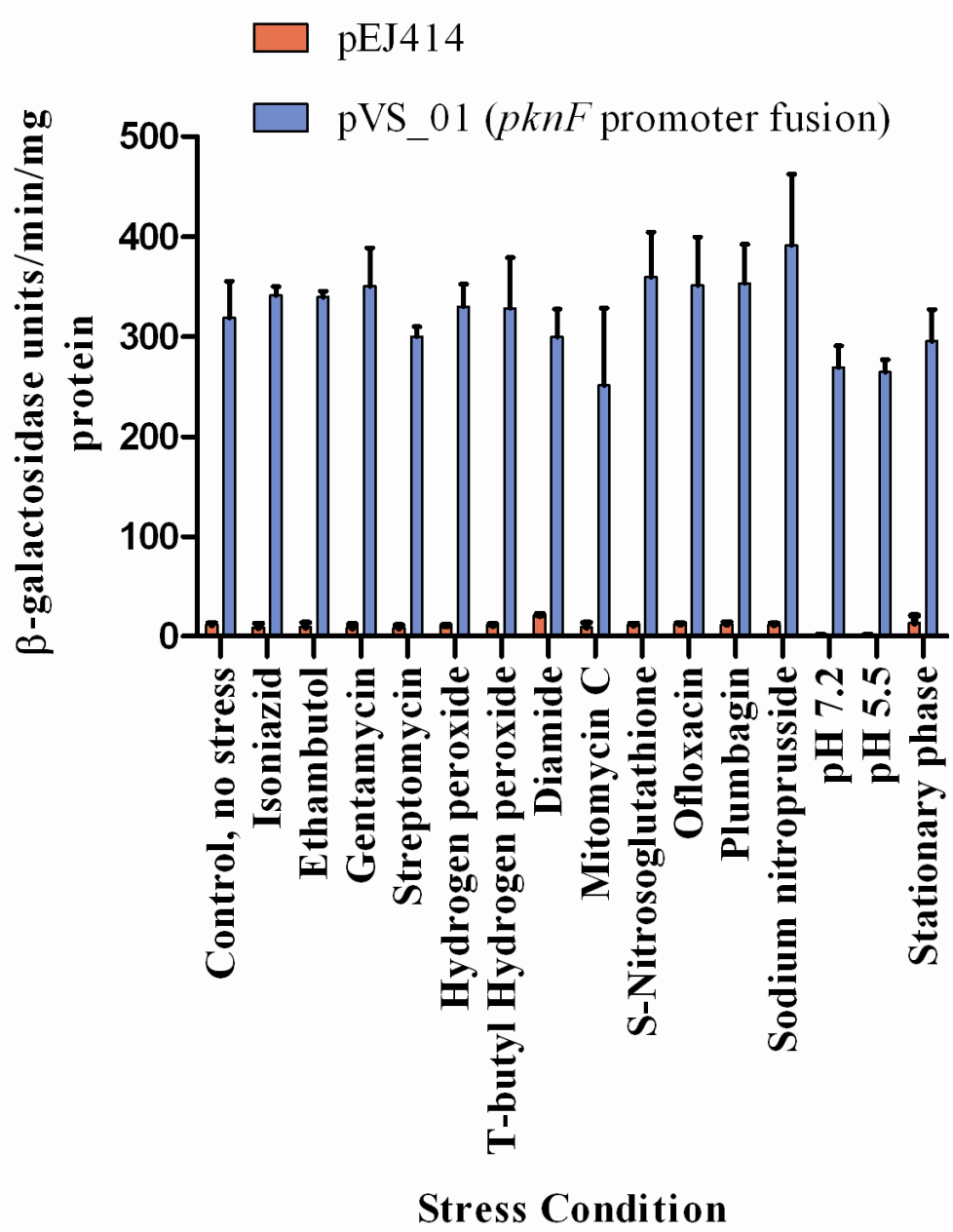


Figure 3.4 ß-galactosidase assays on the *pknF* promoter-*lacZ* strain and pEJ414 control strain in *M. tuberculosis* after a panel of treatments. The ß-galactosidase assay was performed on CFEs of *M. tuberculosis* strains containing pEJ414 and pVS_01 after a panel of treatments. All reagents were added to the *M. tuberculosis* cultures for two hours prior to harvesting for CFE with the exception of acid stress and stationary phase conditions. For acid stress, cultures at an OD_{600} of 0.3 were harvested and resuspended in Dubos adjusted to pH7.2 (control) or Dubos adjusted to pH 5.5 (acid stress) and incubated for a further 24 hours (section 2.4.8). For assessing activity in stationary phase cultures were harvested when at an OD_{600} 2. The data presented are the averages of three biological replicates each assayed in duplicate and the error bars represent standard deviations. In all cases the ß-galactosidase activity levels were significantly higher within strains containing pVS_01 compared with the control strain pEJ414. No other significant differences were noted in the ß-galactosidase activity levels between any of the treatments.

3.4 Assessment of the susceptibility of the *ΔpknF* and *ΔRv1747* mutants to a panel of drug and stress reagents.

Initial characterisation of the *pknF* and *Rv1747* null mutants showed their growth was the same as the parental wild type strain under normal *in vitro* growth conditions (Curry *et al.*, 2005; R. Whalan, personal communication). The next aim of this study was to further characterise both mutants by assessing their susceptibility to a range of drug and stress reagents compared with the wild type H37Rv strain. It was hypothesised that increased sensitivity may be observed when subjecting one or both of the mutant strains to a particular treatment generating clues as to the function of the two proteins in *M. tuberculosis*. To investigate this hypothesis growth inhibition assays were performed using the alamar blue method as outlined in section 2.4.9 and the minimum inhibitory concentrations (MICs) for all three strains subjected to each stress reagent were calculated from the data.

The drugs and stress reagents tested were isoniazid, ethambutol, streptomycin, ciprofloxacin, ofloxacin, hydrogen peroxide, s-nitrosoglutathione, ethidium bromide, mitomycin C and sodium dodecyl sulphate. Ciprofloxacin is another member of the fluoroquinolone family which causes DNA damage by inhibiting DNA gyrase activity necessary for cell division. Sodium dodecyl sulphate (SDS) is an anionic surfactant detergent which can disrupt the cell wall of bacteria and lyse cells in high enough concentrations. It was hypothesised that if one or both of the mutants harboured a defect in their cell wall due to the lack of a particular component that may be transported by Rv1747, then there might be a difference in the MIC between the parental and mutant strains when subjected to SDS treatment. Mycobacteria contain active efflux pumps which are involved in natural drug resistance and ethidium bromide is a small molecule which is known to be pumped out of mycobacterial cells (Danilchanka *et al.*, 2008). For example, Rv0194 is a novel multidrug efflux pump of *M. tuberculosis*; expression of Rv0194 significantly reduced accumulation of ethidium bromide in the cell and conferred multidrug resistance of *Mycobacterium smegmatis* (Danilchanka *et al.*, 2008). Furthermore, Rv1747 belongs to the G subfamily of ABC transporters and ABCG2 has

been identified as a multidrug transporter in humans (Velamakanni et al., 2007). It was therefore hypothesized that the Rv1747 transporter, as a secondary function, might actively export molecules and if Rv1747 was indeed an efflux pump capable of transporting ethidium bromide or one of the fluroquinolones or other antibiotics then there may be an increased sensitivity to one or more of these molecules in the *ΔRv1747* strain. In addition, changes in sensitivity may also be observed in the *ΔpknF* strain depending on the mechanism of regulation between the two proteins (i.e. whether the kinase activity of PknF positively or negatively regulates the transporter function of Rv1747).

All the aforementioned hypotheses were tested in growth inhibition assays and the results of this investigation are shown in figure 3.5. Data were plotted as the concentration of stress reagent versus the percentage growth inhibition. The MICs for all three strains subjected to each stress reagent were calculated from the data and the results are summarized in table 3.1. The MIC was taken to be the lowest concentration of a drug or stress reagent capable of causing ≥90% inhibition compared to the untreated bacteria only controls. **In all cases the *ΔpknF* and *ΔRv1747* mutants were not more susceptible or resistant to any of the drug and stress reagents tested compared with wild type H37Rv**, and hence there were no differences in MICs observed between the strains.

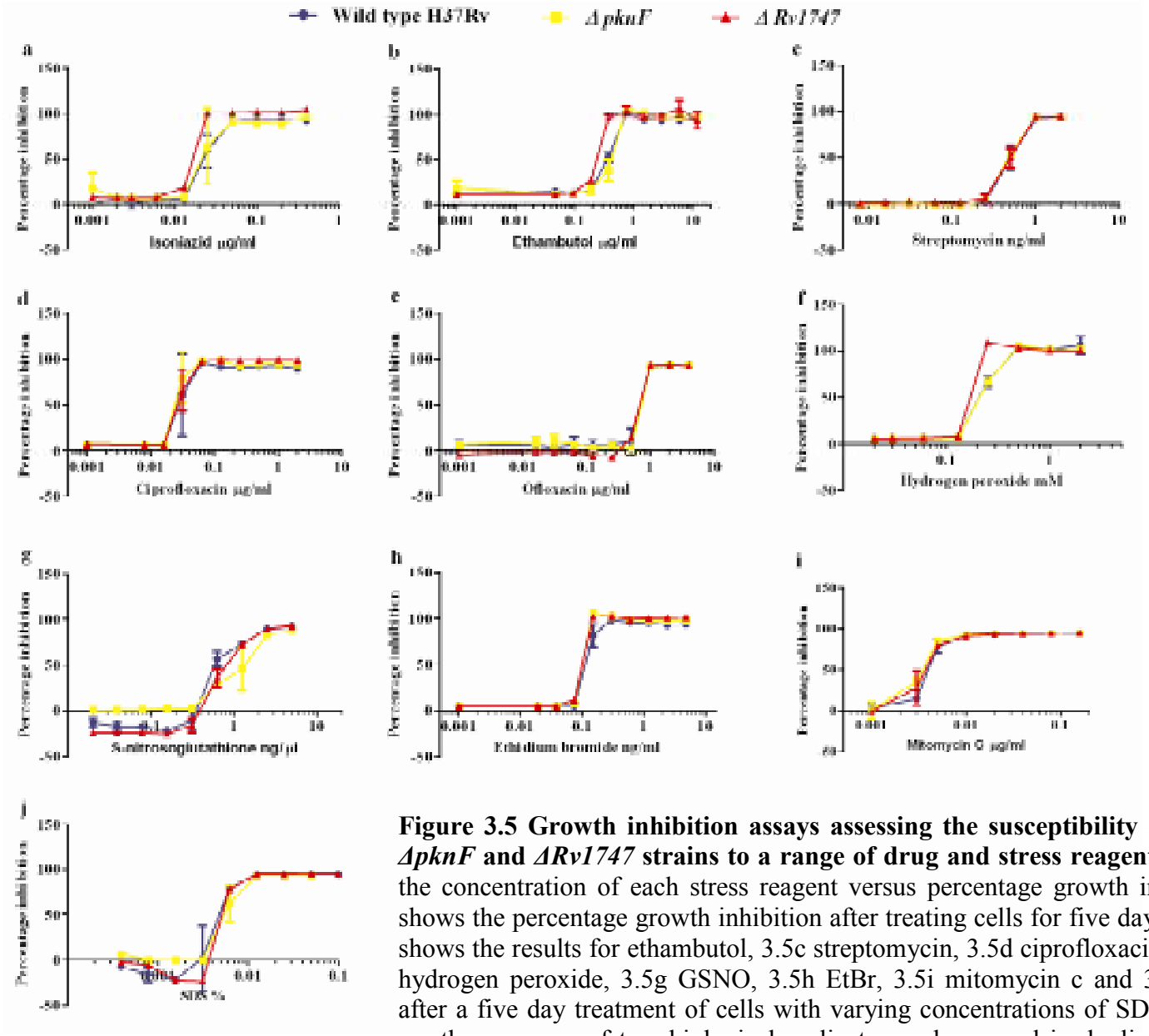


Figure 3.5 Growth inhibition assays assessing the susceptibility of wild type H37Rv, *ΔpknF* and *ΔRv1747* strains to a range of drug and stress reagents. Data are plotted as the concentration of each stress reagent versus percentage growth inhibition. Figure 3.5a shows the percentage growth inhibition after treating cells for five days with isoniazid, 3.5b shows the results for ethambutol, 3.5c streptomycin, 3.5d ciprofloxacin, 3.5e ofloxacin, 3.5f hydrogen peroxide, 3.5g GSNO, 3.5h EtBr, 3.5i mitomycin c and 3.5j shows the results after a five day treatment of cells with varying concentrations of SDS. The data presented are the averages of two biological replicates each assayed in duplicate and the error bars represent standard deviations. In all cases the *ΔpknF* and *ΔRv1747* mutants are not more susceptible or resistant to any of the drug and stress reagents tested compared with wild type H37Rv.

Treatment	**Minimum inhibitory concentration**		
	Wild type	***ΔpknF***	***ΔRv1747***
Isoniazid	0.05 µg/ml	0.05 µg/ml	0.05 µg/ml
Ethambutol	0.75 µg/ml	0.75 µg/ml	0.75 µg/ml
Streptomycin	1 ng/ml	1 ng/ml	1 ng/ml
Ciprofloxacin	0.063 µg/ml	0.063 µg/ml	0.063 µg/ml
Ofloxacin	1 µg/ml	1 µg/ml	1 µg/ml
Hydrogen peroxide	0.5 mM	0.5 mM	0.5 mM
S-nitrosoglutathione	2.5 ng/µl	2.5 ng/µl	2.5 ng/µl
Ethidium bromide	0.15 ng/ml	0.15 ng/ml	0.15 ng/ml
Mitomycin C	10 ng/µl	10 ng/µl	10 ng/µl
SDS	0.025 %	0.025 %	0.025 %

Table 3.1 Summary of the minimum inhibitory concentrations derived from growth inhibition assays of wild type H37Rv, *ΔpknF* and *ΔRv1747* strains after treatment with a panel of drug and stress reagents.

3.5 Discussion

3.5.1 *pknF* and *Rv1747* are co-transcribed

RT-PCR demonstrated that there was a transcript present running from *pknF* through the intergenic region into *Rv1747* showing that these two genes were indeed co-transcribed and formed a single operon. In addition, the presence of *Rv1747* and *pknF* in the same operon gives further merit to the hypothesis that Rv1747 may not only function as an ABC transporter but may also play a role in a signal transduction system with PknF thus warranting further investigation in the following chapters of this study.

However, there is also the possibility that there is a promoter present for *Rv1747* within the coding sequence of *pknF*. For the primer pair that spanned from *pknF* through the intergenic region into *Rv1747* approximately 300 bp of *pknF* transcript and 300 bp of *Rv1747* transcript was amplified. It is possible that there is an *Rv1747* promoter just upstream of the forward primer used in this amplification. Although unlikely, to further investigate this idea one would have to design a probe within *Rv1747* and perform a Northern blot. In addition, one could perform a series of experiments in which constructs of varying lengths were designed upstream of the *Rv1747* translational start site. Each region of DNA would have to be cloned into the *lacZ* fusion vector pEJ414, transformed into *M. tuberculosis* and assessed for levels of promoter activity. If promoter activity was discovered this may also answer the question as to whether there is an *Rv1747* promoter present within the coding region of *pknF*. Additionally 5′RACE or primer extension could be performed to see if a transcriptional start site for *Rv1747* could be mapped. For this study as a whole however, these were not an important subset of experiments to perform and therefore the study progressed into determining the transcriptional start site of *pknF*.

3.5.2 The transcriptional start site of *pknF*

Using 5´ RACE three possible transcriptional start sites were mapped for *pknF*. Due to the caveat that 5´ RACE incorporates a poly-cytosine tail onto the 3´ end of cDNA it was not possible to determine whether transcription started at either of the cytosine residues or the adenine nucleotide which also forms the translation initiation codon (figure 3.2b). It cannot be ruled out, however, that there was secondary structure in the RNA and therefore complete extension to the 5´ end of the message was not achieved. This could be proved using S1 nuclease mapping. Furthermore, by using tobacco pyrophosphatase (TAP) one could distinguish between a processed and unprocessed start site. However, a Solexa sequencing trace of the *M. tuberculosis* transcriptome showed that the mRNA transcript level of the region upstream of the annotated translational start site of *pknF* is negligible and mRNA levels rise sharply immediately after the translation initiation codon suggesting that transcription and translation initiate at the same site (Kristine Arnvig, personal communication).

It was also a possibility that the annotated translational start site of *pknF* was incorrect and this could be tested using translational start site mapping (Smollett *et al.*, 2009). In the *M. tuberculosis* genome, sequencing data suggests that AUG is the preferred start codon with the GUG initiation codon occurring at a lower frequency (Cole *et al.*, 1998) and TTG being a very rare start codon (Salazar *et al.*, 1996). In the case of *pknF* there are 12 amino acids between the annotated start codon according to Tuberculist (http://tuberculist.epfl.ch/) and the beginning of the kinase domain as determined by the SMART domain prediction software (Letunic *et al.*, 2009; Schultz *et al.*, 1998). Within this 12 amino acid sequence there is no alternative start codon present in the genome indicating that the annotated translational start site is correct.

M. tuberculosis promoter consensus sequences are notoriously difficult to identify. A number of studies have been published defining the *M. tuberculosis* sigma factor consensus sequences; a study by Homerova *et al*., in 2008 describes the consensus sequences identified from numerous studies and also identifies a *sigJ* dependent promoter (Homerova *et al.*, 2008). Furthermore, the spacing between the -10 and -35

boxes can vary between 14 bp to 22 bp in length making the search for *M. tuberculosis* promoters all the more challenging. When examining the region upstream of the mapped *pknF* transcriptional and translational start site one finds that it contains a perfect match to the *sigA* dependent -10 consensus sequence of TAyGAT (where y is a C or T residue) 6 nucleotides upstream of the transcriptional start point. In a study examining the distance between transcriptional start sites and the -10 box for *M. tuberculosis* promoters it was found that the distance was most frequently 6 or 7 base pairs (Timm *et al.*, 1999). This provides further evidence supporting the fact that transcription and translation of *pknF* begin at the same nucleotide. Transcription has been shown to initiate at the translational start codon of a few previously studied genes in the mycobacteria. These include the *M. tuberculosis* gene *purC* which is part of the purine biosynthetic pathway (Jackson *et al.*, 1996) and *oxyR* from *M. leprae* which is a DNA binding protein (Dhandayuthapani *et al.*, 1997). Furthermore, in an actinomycete comparative study of transcriptional and translation start sites by Strohl in 1992 they found that there were 11 genes (out of a total of 44 studied) where transcription and translation appeared to start at the same nucleotide suggesting that this is not an uncommon occurrence among the actinomycete family (Strohl, 1992).

3.5.3 Investigation into the transcription of the *pknF-Rv1747* operon after stress

The ß-galactosidase assay results showed that the *pknF* promoter in this experiment had a specific activity of 320 units of ß-galactosidase but there were no significant changes in the levels of ß-galactosidase activity upon treatment with any of the reagents tested. Although there were no changes in the levels of ß-galactosidase activity upon treatment with any of the reagents tested in this study, it does not rule out the possibility that there is a stress condition which alters the level of activity of the *pknF* promoter as was found for *pknE* in the Jayakumar investigation (Jayakumar *et al.*, 2008); it is just that to date the stimulus for *pknF* has not been discovered. In the Jayakumar study the *pknE* promoter was upregulated 1.5 fold after a two hour exposure to the nitric oxide donors sodium nitroprusside and s-nitrosoglutathione. Conversely after treatment with the

oxidative stress agent hydrogen peroxide the promoter activity was decreased after a two hour exposure, although not significantly so (Jayakumar *et al.*, 2008).

The levels of ß-galactosidase activity calculated when testing the effect of acid stress (harvesting cells and resuspending them in Dubos adjusting to pH5.5 for 24 hours) and its control (harvesting cells resuspending them in Dubos at pH7.2 for 24 hours) were lower than all other experiments with an average ß-galactosidase activity of approximately 265 units compared to the no drug or stress control of approximately 320 specific units. The protein content of each CFE sample was determined as outlined in section 2.4.8 to normalise each data set and make the results comparable to one another between the treatments. A possible explanation for the lower ß-galactosidase activity readout for these two sets of data is because some cells may have died when they were washed and re-suspended in fresh medium. Therefore, when normalising to the amount of protein after 24 hours the dead cells would possibly still have contributed to the total amount of protein but obviously there would be no ß-galactosidase activity generated from the plasmid integrated into the dead cells. However, what is important is that the levels of ß-galactosidase activity in the acid stress control and acid stress experiment are the same and therefore tells us that the *pknF* promoter does not respond to low pH/acid stress.

3.5.4 Assessment of the susceptibility of the *ΔpknF* and *ΔRv1747* mutants to a panel of drug and stress reagents.

Growth inhibition assays showed that there were no differences in the susceptibility of the *ΔpknF* and *ΔRv1747* mutants to any of the drug and stress reagents tested compared with wild type H37Rv (figure 3.5), and hence the calculated MICs between the strains for each treatment were the same (see table 3.1). It was hypothesised that increased sensitivity may be observed when subjecting one or both of the mutant strains to a particular treatment generating clues as to the function of the two proteins.

There are many characterised mutants in *M. tuberculosis* which have increased or decreased sensitivity to various reagents; furthermore, this has led to being able to derive functions for these particular proteins. For example as well as the *pknE* mutant

conferring increased resistance to nitric oxide donors (Jayakumar *et al.*, 2008) in 2002 a study showed a knockout of the *M. tuberculosis ahpC* gene which encodes and alkylhydroperoxide reductase showed increased sensitivity to peroxynitrite as did the *M. smegmatis* ortholog knockout (Master *et al.*, 2002). A role was then established for *ahpC* in antioxidant defence mechanisms. Furthermore, in a recent study carried out by Vandal and colleagues in 2009, they discovered four mutants (*Rv2136c*, *Rv2224c*, *ponA2*, and *lysX*) from a screen of over 10, 000 *M. tuberculosis* transposon mutants, that not only were susceptible to acid stress but which were also hypersensitive to SDS, antibiotics and reactive oxygen and nitrogen intermediates demonstrating an association between acid resistance and resistance to other forms of stress (Vandal *et al.*, 2009).

Unfortunately in this present study the *pknF* and *Rv1747* mutants demonstrated no differences in susceptibility to any of the stresses investigated. Yet again this does not mean that there is not a drug or stress reagent which alters the susceptibility of the mutants, it is just that in this piece of research which assessed many of the commonly investigated stress reagents, no reagent was found in which the mutant strains behaved differently to wild type. Furthermore, this study assessed the susceptibility of the two mutants to drug and stress reagents *in vitro*. It is known that growth of the *Rv1747* mutant is impaired in both macrophage and mouse intravenous models of infection (Curry *et al.*, 2005). Therefore another possibility is that one would not see a susceptibility phenotype of these mutants *in vitro* because their function is only required *in vivo*.

In conclusion, **this chapter characterised the *pknF-Rv1747* operon by identifying a transcriptional start site for *pknF* and confirmed co-transcription of the two genes**. An initial assessment of both the mutant strains was then performed but no clues as to the function of the two proteins were gained from these investigations.

CHAPTER 4. INVESTIGATION INTO THE NATURE OF THE Rv1747 SUBSTRATE

4.1 Introduction

Approximately 2.5 percent of the *M. tuberculosis* genome comprises genes encoding putative ABC transporters; a total of 37 ABC transporters have been identified in *M. tuberculosis*, 16 have been categorised as importers and 21 as exporters (Braibant *et al.*, 2000). Rv1747 was classified as a member of a sub-class of transporters which have an unknown function (Braibant *et al.*, 2000). ABC transporters are integral membrane proteins comprising two transmembrane domains and two cytoplasmic nucleotide binding domains. The transporters bind and hydrolyse ATP providing energy for uptake or export of substrates across cell membranes. Functions of ABC transporters include the uptake of nutrients into the cells and the export of virulence factors and toxins. Bacterial ABC importers are typically formed from four polypeptide chains that are often separately encoded (Saurin *et al.*, 1999) and require an external binding protein which functions to deliver the substrate to the transporter (Dawson *et al.*, 2007). In contrast, bacterial exporters are produced as one polypeptide where a single gene usually encodes both the transmembrane domain and an ATPase (nucleotide binding) domain (Saurin *et al.*, 1999). The presence of fused nucleotide binding and transmembrane domains is a strong indicator of an ABC exporter (Davidson *et al.*, 2008; Dawson *et al.*, 2007). Based on its amino acid sequence Rv1747 belongs to the G subfamily of ABC transporters; this family consists of half-transporters which oligomerise to form the functional transporter. Furthermore this protein encodes both the transmembrane and nucleotide binding domains in one polypeptide. Therefore, taken together, from its amino acid sequence and the fact that it is encoded by a single gene Rv1747 is predicted to function as an ABC exporter.

It is known that the function of Rv1747 is required for the growth of *M. tuberculosis* in a mouse intravenous injection model and in bone-marrow derived macrophages and dendritic cells (Curry *et al.*, 2005). Curry *et al*. showed that the *Rv1747* deletion mutant

had a bacterial load which was 10-fold lower than that of the wild type strain in both lungs and spleens of mice and furthermore a similar pattern of growth impairment was observed with the mutant in both macrophages and dendritic cells. This mutant phenotype was lost when a complementing wild type allele of *Rv1747* was introduced into the genome confirming that the growth attenuation observed was due to the deletion of *Rv1747*. What is more, when cultured *in vitro*, the *Rv1747* deletion strain has a similar doubling time to that of the wild type strain illustrating that the function of the Rv1747 protein is only required for the growth of *M. tuberculosis* in cell lines and inside mice, therefore making the function of this protein all the more interesting and relevant to *M. tuberculosis* research.

However, although it has been demonstrated that the function of Rv1747 is required for the growth of the bacterium in mice and macrophages, the identity of the substrate has not yet been revealed. Numerous hypotheses have been raised as to what the nature of the substrate may be by comparing the Rv1747 ABC transporter to analogous systems in other bacteria which are also required for growth and survival within the host. Bacterial ABC exporters transport many different substances which include cell surface components such as lipopolysaccharides, lipids, proteins involved in pathogenesis including haemolysin, and peptides, drugs and siderophores (Dassa & Bouige, 2001); Rv1747 could export any one of these molecules which would make the function of the transporter necessary for growth *in vivo*.

One particular hypothesis raised is that Rv1747 could export a component of the cell wall required for the growth of the bacterium in macrophages and inside mice. In Gram-negative bacteria a multi-protein pathway transports lipopolysaccharide (LPS) across the cell envelope (Ruiz *et al.*, 2009). LPS (also known as endotoxin) is a glycolipid comprised of lipid A, a core oligosaccharide and a long polysaccharide which is present at the outer membrane of Gram-negative bacteria and is responsible for inducing the immune system and causing septic shock. LPS is located at the cell surface and is responsible for making the outer membrane an impermeable barrier which is crucial for the survival of many Gram-negative bacteria in many different environments (Ruiz *et al.*, 2009). The proteins that are required to transport LPS from the inner membrane,

which is the site of synthesis, to the cell surface have been identified. The ABC exporter MsbA flips rough LPS (only containing lipid A and the core region) across the inner membrane. LPS is then extracted from the inner membrane by LptC and three proteins which comprise another ABC exporter (LptF, LptG and LptB). LPS is then transported to an assembly site at the outer membrane (Ruiz *et al.*, 2009). Clearly, LPS transport and assembly is a very complex system which has taken many decades to fully elucidate and comprises of multiple steps and multiple ABC export proteins. Mycobacteria, being gram-positive, do not have LPS, but have other cell wall molecules that may undertake analogous functions with respect to virulence. In *M. tuberculosis* two ABC transporters have been identified which are required for the correct localisation of dimycocerosate (DIM) in the cell envelope (Camacho *et al.*, 2001). These are the DrrABC and MmpL7 transporters. Camacho *et al*. demonstrated that a DIM mutant had increased cell wall permeability and became more sensitive to SDS which is associated with an increase in the outer membrane permeability (Camacho *et al.*, 2001) although the DIM mutant had similar growth rate to the wild type in naïve macrophages (Camacho *et al.*, 1999). In analogy with LPS in gram-negative bacteria, it is hypothesised that Rv1747 may function to export a protein or lipid component of the *M. tuberculosis* outer membrane that is also necessary for growth and survival within the host.

Another hypothesis to be considered in this study was that Rv1747 may be exporting a protein required for growth. For example Rv1747 could be involved in exporting a siderophore involved in iron uptake. Siderophores are small compounds secreted by bacteria which act to chelate iron; iron is essential for many fundamental cellular processes and is transported back inside the mycobacterial cell. In *M. tuberculosis* one iron siderophore exporter has already been characterised; the ABC transporter IrtA (Rv1348), which contains a substrate binding domain and ATPase domain, is a siderophore exporter whilst the IrtB-Rv2895c system functions to import ferrated siderophores (Farhana *et al.*, 2008). IrtB is an ABC transporter and Rv2895c is a siderophore binding protein (Farhana *et al.*, 2008).

It has been suggested (http://tuberculist.epfl.ch/) that the substrate of the Rv1747 ABC transporter may be a lipooligosaccharide. A review of all the *M. tuberculosis* ABC

transporters published in 2000 revealed that Rv1747 falls into a sub-class of transporters which have an unknown function (Braibant *et al.*, 2000). Rv1747 is highly similar to the White protein from *Drosophila melanogaster* which is a permease necessary for the transport of pigment precursors responsible for eye colour and NodI from *Rhizobium* strains which is a protein implicated in the nodulation process by export of a polysaccharide (Braibant *et al.*, 2000). I have compared the homology with other known ABC transporters from other organisms, and in addition to the two proteins mentioned above, an NCBI Blast of the Rv1747 protein revealed similarities with a cobalt transport family protein found in bacteria, eukaryotes and archaea and with an ABC group A subfamily which functions to transport a variety of lipid compounds suggesting that the Rv1747 transporter could be involved in one of these functions in *M. tuberculosis*. Another hypothesis is that Rv1747 could be exporting a small molecule. As previously mentioned the Rv1747 transporter shares similarity with a cobalt transport family protein. Cobalt is a transition metal which is an essential component of many enzymes.

It must also be taken into consideration that there is no conclusive evidence to prove that Rv1747 does not function as an ABC importer; however all the gathered evidence suggests that Rv1747 most likely functions as an exporter. Consequently, the experiments performed in this study to search for the Rv1747 substrate were designed in such a way so that they were not dependent on knowing the direction of substrate transport. All the aforementioned hypotheses will be explored and discussed in this chapter of the study with the aim to identify the Rv1747 substrate which is required for growth of the bacterium *in vivo*.

4.2 Transcriptional microarray studies of the *ΔpknF* and *ΔRv1747* mutant strains

As the identity of the Rv1747 substrate is unknown the initial approach taken to generate a hypothesis as to the nature of the substrate was to use cDNA microarray technology to analyse the effects of *M. tuberculosis* global gene expression in the *pknF* and *Rv1747* deletion strains. Microarrays can be used as a screening tool to allow for comparative analysis of the entire transcriptome of all genes in an organism simultaneously.

Transcriptional microarrays were used in this study to compare gene expression between wild type H37Rv and both the *pknF* and *Rv1747* deletion strains. Changes in gene expression can allow one to obtain insights into the possible function of proteins.

Global gene expression differences were evaluated between wild type *M. tuberculosis* and both the *pknF* and *Rv1747* deletion strains using RNA-DNA microarrays. RNA from wild type *M. tuberculosis* and RNA extracted from a null mutant were directly compared; labelled cDNAs were competitively hybridised against a single array. Three biological replicates were performed for each condition and they were carried out in technical duplicate (dye-swaps) as described in section 2.4.5 and Appendix III. A t-test against zero was performed for each data-set. Data were taken to be significant if the p-value was less than 0.05 and a two-fold level of change cut-off was initially applied to all the data (Appendix III). In order to be certain that the transcriptional profiles observed were due to the *Rv1747* and *pknF* deletions microarrays were also performed comparing gene expression levels in wild type H37Rv and both the complementing strains. The genes present in all the Results Tables only include those whose differential gene regulation was restored to wild type levels in the complementing strains. Seven genes in the *Rv1747* deletion experiment and three genes in the *pknF* deletion experiment were not complemented and are listed in tables 4.1 and 4.2 respectively. As these genes were not complemented upon replacement of the *Rv1747* or *pknF* genes then they were considered not important to this investigation as they would not help to generate clues as to the function of the two proteins.

4.2.1 The expression of many *M. tuberculosis* genes were affected by *Rv1747* deletion

When applying a two-fold level of change cut-off to the data comparing global gene expression in wild type H37Rv and *ΔRv1747* strains there were 40 genes whose relative transcription level differed significantly upon *Rv1747* deletion and whose differential expression was restored in the complementing strain (see Appendix IV). Fourteen of these genes were up-regulated and 26 were down-regulated. The microarray data for the

ten most highly up-regulated and down-regulated genes within the *Rv1747* deletion strain are shown in tables 4.3 and 4.4 respectively. When the fold change cut-off level

Gene Number	Gene Name	Gene Product	Fold change in mutant	Fold change in complement
Rv0281	*Rv0281*	Conserved hypothetical protein	2.0 (up-reg)	2.9 (up-reg)
Rv0282	*eccA$_3$*	Esx conserved component	2.1 (up-reg)	1.9 (up-reg)
Rv0796	*Rv0796*	Putative transposase	2.2 (down-reg)	1.5 (down-reg)
Rv1764	*Rv1764*	Putative transposase	2.2 (down-reg)	1.5 (down-reg)
Rv2031c	*HspX*	Heat shock protein	2.6 (up-reg)	2.5 (up-reg)
Rv2106	*Rv2106*	Probable tranposase	2.1 (down-reg)	1.4 (down-reg)
Rv3131	*Rv3131*	Conserved hypothetical protein	2.1 (up-reg)	6.32 (down-reg)

Table 4.1 Microarray data for the genes whose differential expression levels were not restored upon complementation with the *Rv1747* gene. Up-reg = fold up-regulated. Down-reg = fold down-regulated.

Gene Number	Gene Name	Gene Product	Fold change in mutant	Fold change in complement
Rv3614c	*espD*	Esx1 secretion associated protein	2.5 (down-reg)	2.5 (down-reg)
Rv3615c	*espC*	Esx1 secretion associated protein	2.3 (down-reg)	2.4 (down-reg)
Rv3616c	*espA*	Esx1 secretion associated protein	2.1 (down-reg)	2.3 (down-reg)

Table 4.2 Microarray data for the genes whose differential expression levels were not restored upon complementation with the *pknF* gene. Up-reg = up-regulated. Down-reg = down regulated.

was reduced to 1.5 there were 483 genes whose transcription was differentially regulated between the wild type and *Rv1747* deletion strain which is approximately 5% of the *M. tuberculosis* genome.

Expression of *Rv1747* was 29 fold down-regulated in the *ΔRv1747* strain compared to wild type which was expected as *Rv1747* is the gene which was deleted from the genome (table 4.4). In the *Rv1747* complement strain expression of *Rv1747* was 1.3 fold up-regulated compared to wild type confirming the restoration of gene expression.

The gene most up-regulated in the *ΔRv1747* strain was the isoniazid-inducible gene *iniB* (table 4.1). This gene was 3.8 fold up-regulated in the null mutant compared to wild type with a p-value of 2.4 $x10^{-4}$. Furthermore the second most upregulated gene was *iniA* (3.2 fold, p-value of $8.14x10^{-5}$). *iniB* forms an operon with *iniA* and *iniC*. Interestingly *iniC* was also upregulated 1.6 fold (p-value of $6.48x10^{-4}$). The promoter of the *iniBAC* operon is specifically induced by a broad range of inhibitors to cell wall biosynthesis (Alland *et al.*, 2000). The list of the top ten up-regulated genes also included the probable acyl-CoA dehydrogenase *fadE23* (2.1 fold change). A probable ferredoxin, *fdxA*, was also up-regulated 2.1 fold. Genes with a 2.0 fold up-regulation in the null mutant also comprised *gyrB* (DNA gyrase subunit B), PE_PGRS41 (a PE_PGRS family protein), *ethA* (whose gene product activates the pro-drug ethionamide) and *ino1* (involved in the phosphatidyl-myo-inositol (PI) biosynthetic pathway). The two conserved hypothetical proteins *Rv0047c* (2.2 fold) and *Rv0822c* (2.0 fold) were also included in the list. A search of the *M. tuberculosis* kinases revealed that *pknF* was up-regulated 1.9 fold (p-value 0.001) in the *ΔRv1747* strain as was *pknD* (1.8 fold, p-value 0.001). Interestingly *pknF* expression was still 2.0 fold up-regulated with a p-value of 0.003 in the *Rv1747* complement strain.

Gene Number	Gene Name	Gene Product	Fold Up-regulated	p-value
Rv0341	*iniB*	Isoniazid inducible gene protein IniB	3.8	2.4E-04
Rv0342	*iniA*	Isoniazid inducible gene protein IniA	3.2	8.1E-05
Rv0047c	*Rv0047c*	Conserved hypothetical protein	2.2	4.1E-05
Rv3140	*fadE23*	Probable acyl-CoA dehydrogenase FadE23	2.1	0.001
Rv2007c	*fdxA*	Probable ferredoxin FdxA	2.1	0.005
Rv0822c	*Rv0822c*	Conserved hypothetical protein	2.0	4.E-05
Rv0005	*gyrB*	DNA gyrase subunit B	2.0	0.001
Rv2396	*PE_PGRS41*	PE-PGRS family protein	2.0	7.E-04
Rv3854c	*ethA*	Monooxygenase EthA	2.0	0.003
Rv0046c	*ino1*	Myo-inositol-1-phosphate synthase Ino1	2.0	3.E-04

Table 4.3 Microarray data for the ten *M. tuberculosis* genes most highly up-regulated upon *Rv1747* deletion

Gene Number	Gene Name	Gene Product	Fold Down-regulated	p-value
Rv1747	*Rv1747*	Probable conserved transmembrane ATP-binding protein ABC transporter	29.2	8.9E-06
Rv2415c	*Rv2415c*	Conserved hypothetical protein	2.3	7.9E-04
Rv1380	*pyrB*	Probable aspartate carbamoyltransferase PyrB	2.2	0.001
Rv1999c	*Rv1999c*	Probable conserved integral membrane protein	2.2	0.004
Rv2814c	*Rv2814c*	Probable transposase	2.2	0.011
Rv1382	*Rv1382*	Probable export or membrane protein	2.2	6.2E-04
Rv2528c	*mrr*	Probable restriction system protein Mrr	2.1	1.9E-04
Rv1040c	*PE8*	PE family protein	2.1	0.013
Rv2265	*Rv2265*	Possible conserved integral membrane protein	2.1	0.005
Rv2577	*Rv2577*	Conserved hypothetical protein	2.1	0.015

Table 4.4 Microarray data for the ten *M. tuberculosis* genes most highly down-regulated upon *Rv1747* deletion.

Apart from *Rv1747*, other down-regulated genes in the top ten list included the conserved hypothetical proteins *Rv2415c* (2.3 fold) and *Rv2577* (2.1 fold) (table 4.4). Two possible conserved integral membrane proteins, namely *Rv1999c* and *Rv2265*, were also down-regulated in the null strain 2.2 and 2.1 fold respectively. *mrr* which is involved in the acceptance of foreign DNA was down-regulated 2.1 fold. The probable export/membrane protein Rv1382 was also downregulated 2.2 fold. Finally the probable aspartate carbamoyltransferase *pyrB* was down-regulated 2.2 fold as was *Rv2814c*, a probable transposase.

4.2.2 The expression of only a few *M. tuberculosis* genes were affected by *pknF* deletion

Upon examination of the transcriptional profiles of the arrays which hybridised the wild type and *ΔpknF* strains, when a two fold cut-off was applied to the data the relative transcription level of only 12 genes were differentially regulated greater or equal to two-fold. In addition all 12 genes were found to be down-regulated. Due to the low number of differentially regulated genes passing a two fold cut-off the stringency was reduced to 1.5 fold. With a 1.5 fold cut-off the number of genes found to be differentially regulated rose to 72 and they are listed in Appendix IV. Of the 72 differentially expressed genes only 12 were up-regulated. The microarray data for the ten most highly up-regulated and down-regulated genes within the *pknF* deletion strain are shown in tables 4.5 and 4.6 respectively. The *pknF* gene itself did not appear in the microarray results list. This was because the gene did not pass the filtering stages within the microarray analysis.

Interestingly, the gene most upregulated in the *ΔpknF* strain was *iniB* (1.8 fold, p-value 0.001), the same gene that was most highly up-regulated in the *ΔRv1747* strain. *iniA* was also upregulated 1.5 fold in the *pknF* deletion strain with a p-value of 0.005. Other genes present in the *ΔpknF* list that were also up-regulated in the *Rv1747* null strain were *fdxA* (1.8 fold) and *ethA* (1.7 fold up-regulated). *glpQ1*, a probable glycerophosphoryl diester phosphodiesterase, was up-regulated 1.8 fold in the *pknF* null mutant. The conserved

hypothetical proteins *Rv3850* and *Rv1738* were up-regulated 1.7 and 1.6 fold respectively in the *ΔpknF* strain. *espE*, an esx1 secretion-associated protein

Gene Number	**Gene Name**	**Gene Product**	**Fold Up-regulated**	**p-value**
Rv3842c	*glpQ1*	Probable glycerophosphoryl diester phosphodiesterase GlpQ1	1.8	0.006
Rv2007c	*fdxA*	Probable ferredoxin fdxA	1.8	0.023
Rv0341	*iniB*	Isoniazid inductible gene protein iniB	1.8	0.001
Rv3854c	*ethA*	Monooxygenase EthA	1.7	0.006
Rv3850	*Rv3850*	Conserved hypothetical protein	1.7	0.002
Rv1738	*Rv1738*	Conserved hypothetical protein	1.6	0.005
Rv3864	*espE*	Esx-1 secretion-associated protein EspE	1.6	0.003
Rv3727	*Rv3727*	Possible oxidoreducatse	1.5	0.008
Rv0175	*Rv0175*	Probable conserved mce associated membrane protein	1.5	0.002
Rv3728	*Rv3728*	Probable conserved two domain membrane protein	1.5	0.006

Table 4.5 Microarray data for the ten *M. tuberculosis* genes most highly up-regulated upon *pknF* deletion

Gene Number	Gene Name	Gene Product	Fold Down-regulated	p-value
Rv2480c	*Rv2480c*	Possible transposase for insertion sequence element IS6110	2.4	0.003
Rv1371	*Rv1371*	Probable conserved membrane protein	2.4	0.042
Rv2106	*Rv2106*	Probable transposase	2.0	0.029
Rv1372	*Rv1372*	Conserved hypothetical protein	2.2	0.037
Rv1370c	*Rv1370c*	Putative transposase for insertion sequence element IS6110	2.0	0.042
Rv3640c	*Rv3640c*	Probable transposase	2.0	0.036
Rv2815c	*Rv2815c*	Probable transposase	2.0	0.044
Rv2515c	*Rv2515c*	Conserved hypothetical protein	2.0	0.034
Rv2167c	*Rv2167c*	Probable transposase	2.0	0.015
Rv0796	*Rv0796*	Putative transposase for insertion sequence element IS6110	2.0	0.035

Table 4.6 Microarray data for the ten *M. tuberculosis* genes most highly down-regulated upon *pknF* deletion

was also up-regulated 1.6 fold. *Rv3727* (probably involved in cellular metabolism), *Rv3728* (probably involved in an efflux system) and *Rv0175* (a probable conserved mce associated membrane protein) were all up-regulated 1.5 fold.

The top ten list of genes down-regulated in the *ΔpknF* strain comprised seven possible transposases; *Rv2480c* (2.4 fold), and *Rv1380*, *Rv2106*, *Rv3640c*, *Rc2815c*, *Rv2167c*, *Rv0796* (all 2.0 fold down-regulated). The remaining three differentially regulated genes were *Rv1371* (2.4 fold) and *Rv1372* (2.2 fold) which are annotated as a probable conserved membrane protein and a conserved hypothetical protein and *Rv2515c* (2.0 fold), also annotated as a conserved hypothetical protein.

4.2.3 The effect of the *pknF* and *Rv1747* deletions were confirmed for a selection of genes

To confirm the effect of the *pknF* and *Rv1747* deletions on gene expression, a selection of genes were tested for their altered relative abundance of transcripts in the null strains using quantitative real-time PCR (qRT-PCR). It was particularly important to determine the levels of expression of *pknF* in the *pknF* null and complement strain as this gene did not pass the filtering within the microarray analysis. qRT-PCR was performed as outlined in section 2.3.3 and a list of primers can be found in Appendix II. qRT-PCR was initially performed with genomic DNA to check that a good amplification plot was achieved for each primer pair and to confirm the absence of non-specific products (data not shown). qRT-PCR was then performed with cDNA made from RNA extracts from all five strains with *sigA* primers (the gene to which all results were to be normalised) to confirm that the expression of this gene did not alter between the five strains. Results are shown in Figures 4.1a (wild type, *pknF* null and *pknF* complement strains) and 4.1b (wild type, *Rv1747* null and *Rv1747* complement strains) and demonstrate that there were no differences in the expression of *sigA* between the strains.

Figure 4.2 shows the results of qRT-PCR for wild type, *ΔRv1747* and the respective complement strain. Data are the means of three biological replicates and the error bars show the standard deviations. The normalised transcription level for the following genes were plotted; *Rv1747* (figure 4.2a), *pknF* (figure 4.2b), *iniB* (figure 4.2c), *iniA* (figure 4.2d), *iniC* (figure 4.2e), *ethA* (figure 4.2f) and *pknD* (figure 4.2g). Figure 4.2a confirms that expression of *Rv1747* in the null mutant is almost undetectable and transcription is restored to wild type levels in the complementing strain. Figure 4.2b shows that the levels of *pknF* transcript do not increase in the *ΔRv1747* strain but do increase in the complementing strain, presumably due to the presence of an extra copy of *pknF* in the *Rv1747* complementing plasmid. Microarray results also showed that *pknF* was up-regulated in the *Rv1747* complementing strain but unlike the microarray data the level of *pknF* transcript does not increase in the *ΔRv1747* strain when determined by qRT-PCR. Figures 4.2 c-e show that the transcriptional profiles of these genes follows the same patterns as shown by the microarray results; transcript abundance levels all increase in the *Rv1747* mutant strain and are complemented when the *Rv1747* gene was replaced. This is particularly striking for *iniB* (figure 4.2c) and *iniA* (figure 4.2d) where there is approximately three times as much transcript present in the mutant strain compared with wild type. These results confirm the microarray results which showed that *iniB* and *iniA* were up-regulated by 3.8 and 3.2 fold respectively in the *ΔRv1747* strain.

Figure 4.3 shows the results of qRT-PCR for wild type, *ΔpknF* and the respective complemented strain. Again, data are the means of three biological replicates and the error bars show the standard deviations. The normalised transcription level for the following genes were plotted; *pknF* (figure 4.3a), *Rv1747* (figure 4.3b), *iniB* (figure 4.3c), *iniA* (figure 4.3d) and *iniC* (figure 4.3e). Figure 4.3a confirms that expression of *pknF* in the null mutant is undetectable and transcription is restored to almost wild type levels in the complementing strain. Figure 4.3b shows that the levels of *Rv1747* transcript are the same in all three strains. Figures 4.3 c-e show that the transcriptional profiles of *iniB* (figure 4.3c), *iniA* (figure 4.3d) and *iniC* (figure 4.3e) follow the same pattern as shown by the microarray data (i.e. *iniB* and *iniA* have increased normalised transcript levels in the *ΔpknF* strain which is restored to wild type levels in the complement). The *iniC* gene did not appear in the gene lists generated by microarrays in

the *pknF* mutant or complement strain. Indeed, figure 4.3e clearly show that the levels of *iniC* transcript do not change in the *pknF* mutant strain but are in fact slightly decreased in the *pknF* complemented strain.

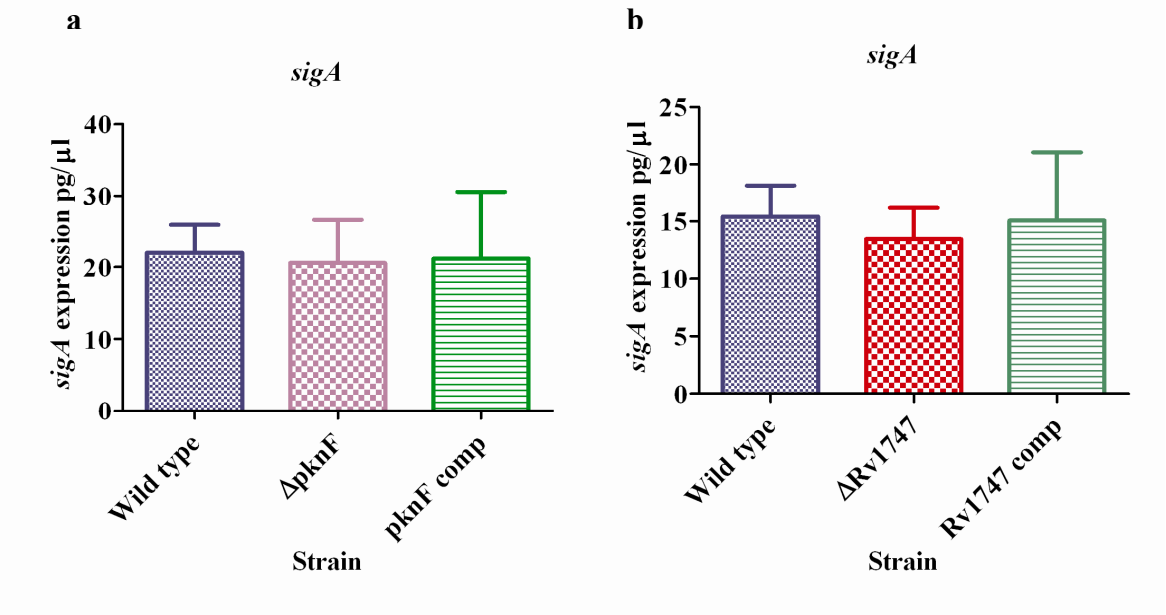


Figure 4.1 qRT-PCR to confirm that transcription of *sigA* does not change between wild type, *ΔpknF*, *pknF* complement, *ΔRv1747* and *Rv1747* complement strains. qRT-PCR was performed with cDNA which was made as described in 2.3.3. Figure a shows the expression of *sigA* in wild type, *ΔpknF* and *pknF* complement strains. Figure b shows the expression of *sigA* in wild type, *ΔRv1747* and *Rv1747* complement strains. Data plotted are the mean of three biological replicates and the error bars show the standard deviations. Data show there were no significant differences in levels of *sigA* expression between the strains.

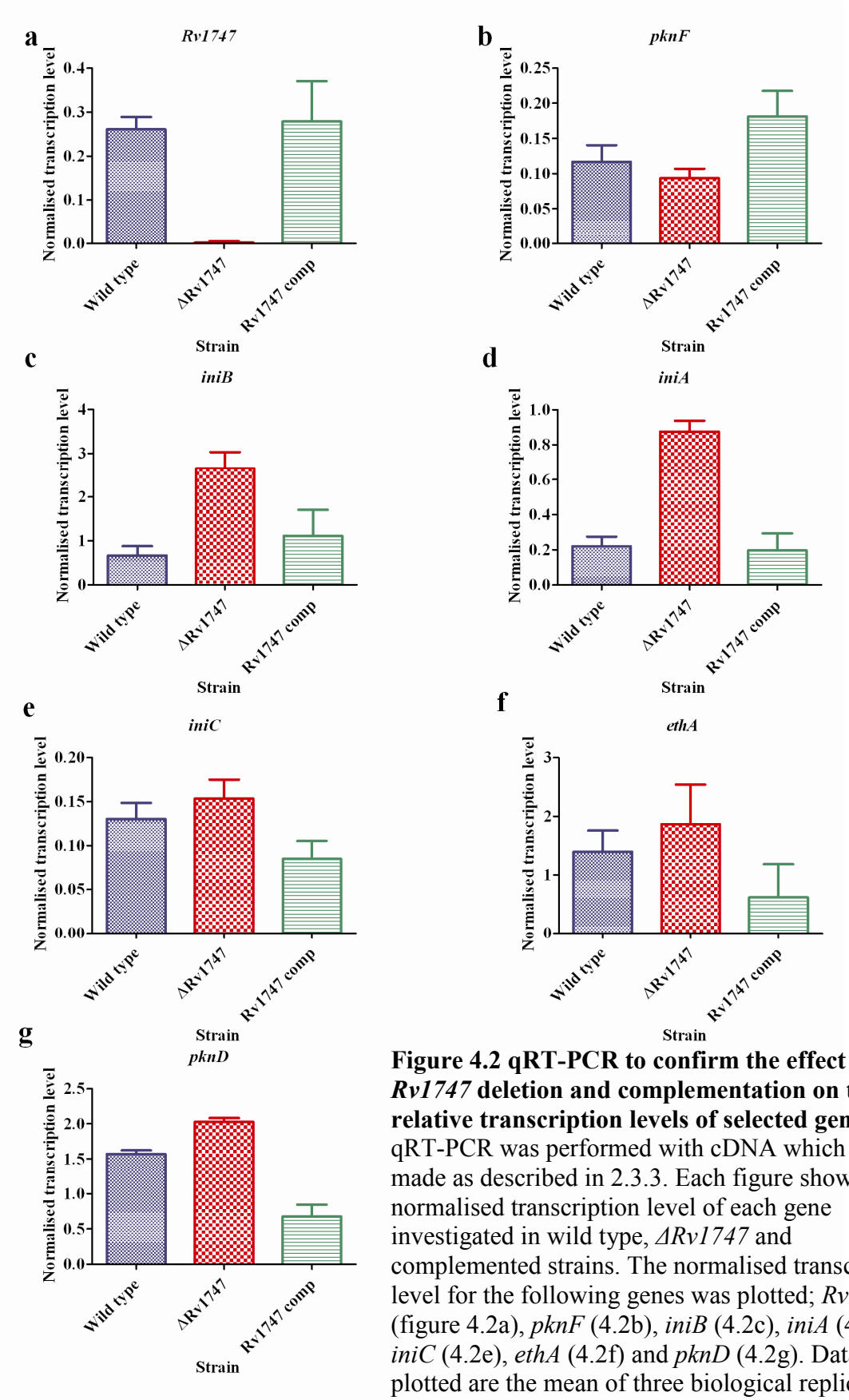


Figure 4.2 qRT-PCR to confirm the effect of *Rv1747* deletion and complementation on the relative transcription levels of selected genes. qRT-PCR was performed with cDNA which was made as described in 2.3.3. Each figure shows the normalised transcription level of each gene investigated in wild type, *ΔRv1747* and complemented strains. The normalised transcription level for the following genes was plotted; *Rv1747* (figure 4.2a), *pknF* (4.2b), *iniB* (4.2c), *iniA* (4.2d), *iniC* (4.2e), *ethA* (4.2f) and *pknD* (4.2g). Data plotted are the mean of three biological replicates and the error bars show the standard deviations.

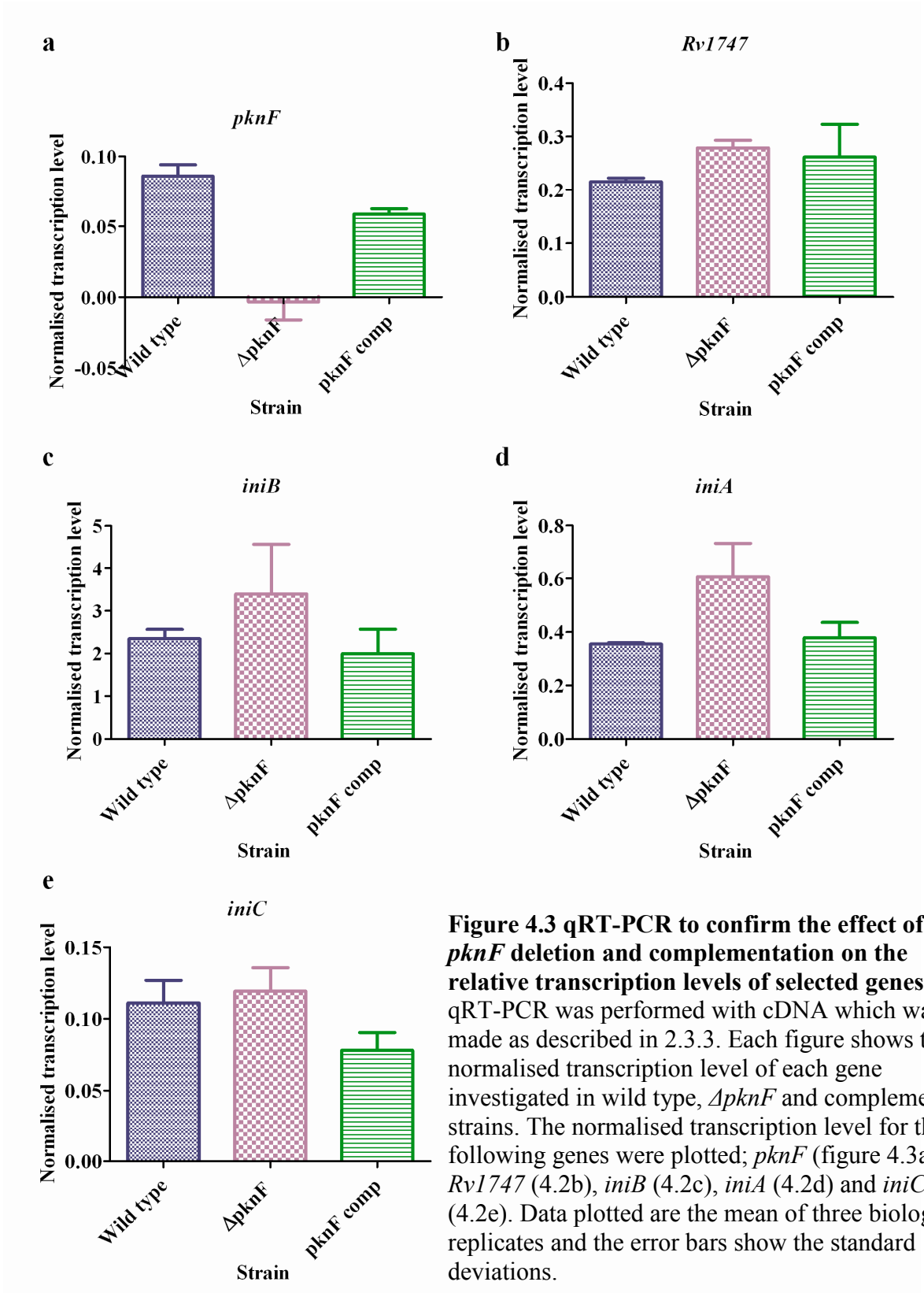


Figure 4.3 qRT-PCR to confirm the effect of *pknF* deletion and complementation on the relative transcription levels of selected genes. qRT-PCR was performed with cDNA which was made as described in 2.3.3. Each figure shows the normalised transcription level of each gene investigated in wild type, *ΔpknF* and complement strains. The normalised transcription level for the following genes were plotted; *pknF* (figure 4.3a), *Rv1747* (4.2b), *iniB* (4.2c), *iniA* (4.2d) and *iniC* (4.2e). Data plotted are the mean of three biological replicates and the error bars show the standard deviations.

4.3 Assessment of the lipid composition of the *M. tuberculosis* cell wall

Microarray and qRT-PCR analysis (section 4.2) showed there to be up-regulation of genes involved in cell wall stress and synthesis in both of the null mutants. Therefore it was hypothesised that Rv1747 may export a component of the cell envelope. To further investigate this hypothesis two dimensional thin-layer chromatography (2D TLC) was used to analyse the total lipid composition of H37Rv wild type, *ΔRv1747* and *ΔpknF* strains plus their respective complements. The *pknF* mutant was included in this study for two reasons; firstly, the differentially regulated gene lists generated in the microarray studies were similar for both the *ΔpknF* and *ΔRv1747* mutants and secondly, PknF is hypothesised to regulate the transporter activity of Rv1747. Therefore it made sense to include this strain plus its respective complement in the study because depending on whether phosphorylation of Rv1747 by PknF positively or negatively regulates the transporter activity of the protein then we may have observed similar differences, if there were any, in the lipid content of the cell in both the kinase and transporter mutants.

Cultures of wild type H37Rv, *ΔRv1747*, *ΔpknF* and their respective complements were grown in Dubos medium to exponential phase (OD_{600} 0.6-0.8) and then labelled with ^{14}C-acetate for 24 hours as described in section 2.4.6. Acetate is a carbon source for lipid synthesis and therefore was incorporated into all newly synthesised lipids over the 24 hour period. *M. tuberculosis* has a doubling time of approximately 18 hours so by labelling for 24 hours it ensured that a full round of replication had occurred. All lipids were then extracted both from the cell pellets and from the culture supernatants for analysis by 2D TLC.

Figure 4.4 shows the initial panel of 2D TLC experiments that were performed examining the lipid profiles of the cell pellets in wild type and both mutant strains. Systems A to D show separation of the apolar lipids and systems D and E show separation of the polar lipids. See table 2.4 in the Material and Methods section for the list of the combinations of solvents used in each system. System A separated triacyl glycerol (TAG) and phthiocerol dimycocerosates (PDIM). System B separated diacyl glycerol (DAG), free fatty acids and free mycolic acids. System C separated DAG and free fatty acids. Apolar system D

separated diacyl trehalose (DAT), glucose monomycolate (GMM), sulpholipids 1, 2 and 3 (SL-III), phospholipids (P) and trehalose dimycolates (TDM). Polar system D separated DAT and system E separated P, phosphatidyl-myo-inositol (PI) and acylated phosphatidyl-myo-inositol dimannoside (AC_2PIM_2). Upon examination of the profiles of the polar lipids in system D it was found that there was an accumulating lipid in the *pknF* and *Rv1747* mutant strains. Blue arrows highlight this lipid in figure 4.4. The lipid is an unidentified lipid which had not been previously characterised.

The lipid profiles of the complementing strains were then analysed. Results showing the polar lipids present in system D for all five strains are in figure 4.5. The chromatography shows that the accumulating lipid present in both the knockout strains, again highlighted with arrows, is also present in both of the complemented strains demonstrating that the lipid difference was not restored when either gene was replaced. Next, to determine whether the lipid accumulating within the cell of the mutants (and complementing strains) were a result of an inability to transport the lipid across the cell membrane, the profiles of the lipids present in the culture supernatants were analysed. Figure 4.6 shows the results of the chromatography examining the polar lipid profiles of the culture supernatants after separation in polar system D, the same system which showed the previous accumulation. Figure 4.6 shows that there is in fact a lipid present in the wild type strain of the culture supernatants (figure 4.6a) that is not present in either of the deletion or complementing strains (figures 4.6b to 4.6e). It is not possible to say whether this lipid present in the culture supernatants is the same one as observed in the cell pellets of the mutant and complementing strains.

In summary, the results showed that in the cell pellets there was a lipid accumulating in the mutant strains in system D (figure 4.5) and in the culture supernatants there was a lipid present in the wild type strain, in the same system, that was absent in the deletion strains (figure 4.6). However, the reciprocal lipid accumulations were not restored in the complemented strains.

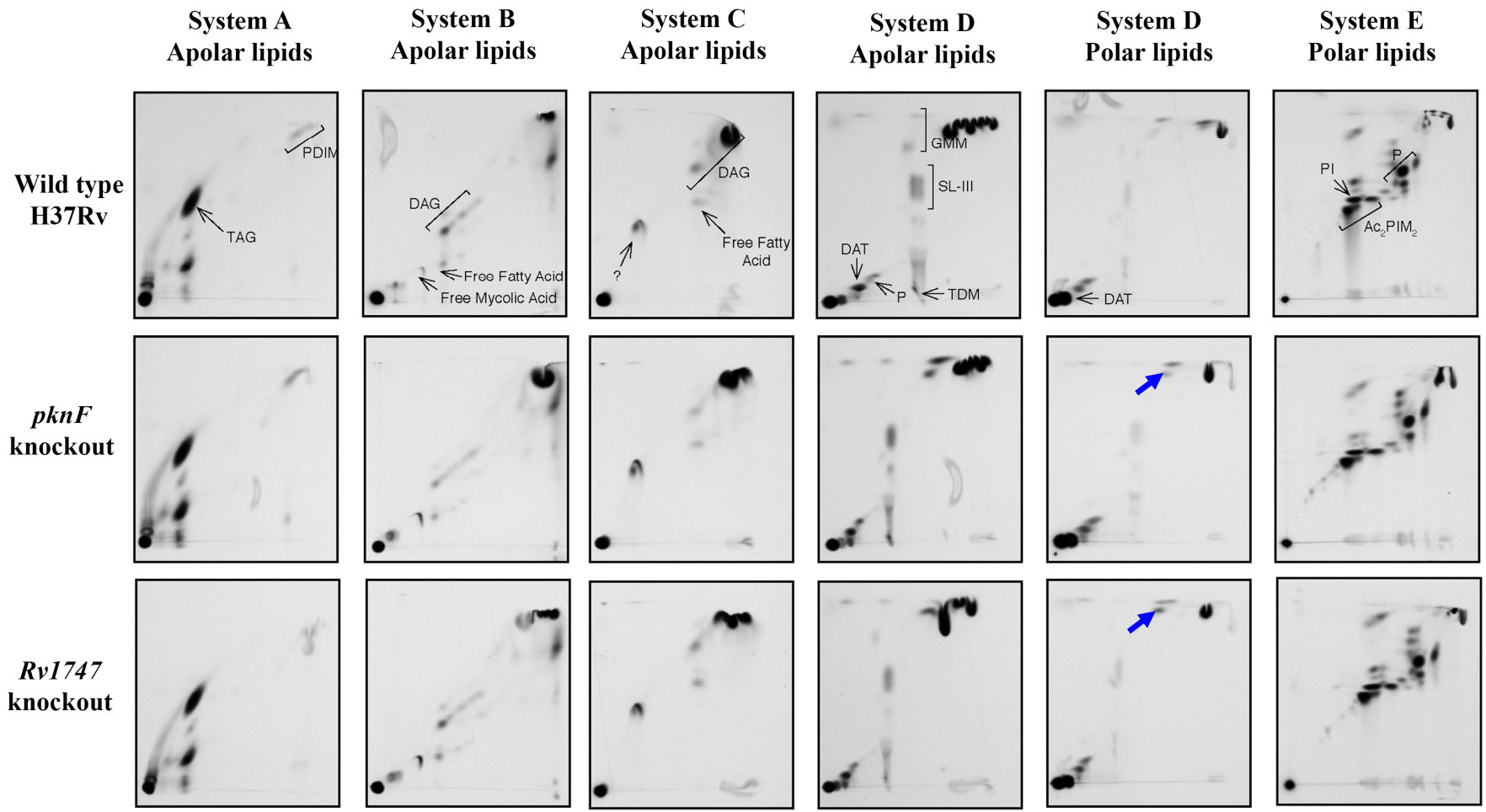


Figure 4.4 Two dimensional thin-layer chromatography of wild type H37Rv, *ΔpknF* and *ΔRv1747* strains showing the apolar and polar lipids present in the cell pellets after separation in five different systems. Lipids were extracted as outlined in section 2.4.6. 20, 000 cpm of each sample was spotted onto a silica plate and then the lipids were separated in two different directions before being exposed to x-ray film. The top panels label each known and characterized lipid. For information about the acronyms see text in Results section 4.3. The blue arrows highlight the accumulating polar lipid in the *ΔpknF* and *ΔRv1747* strains in system D.

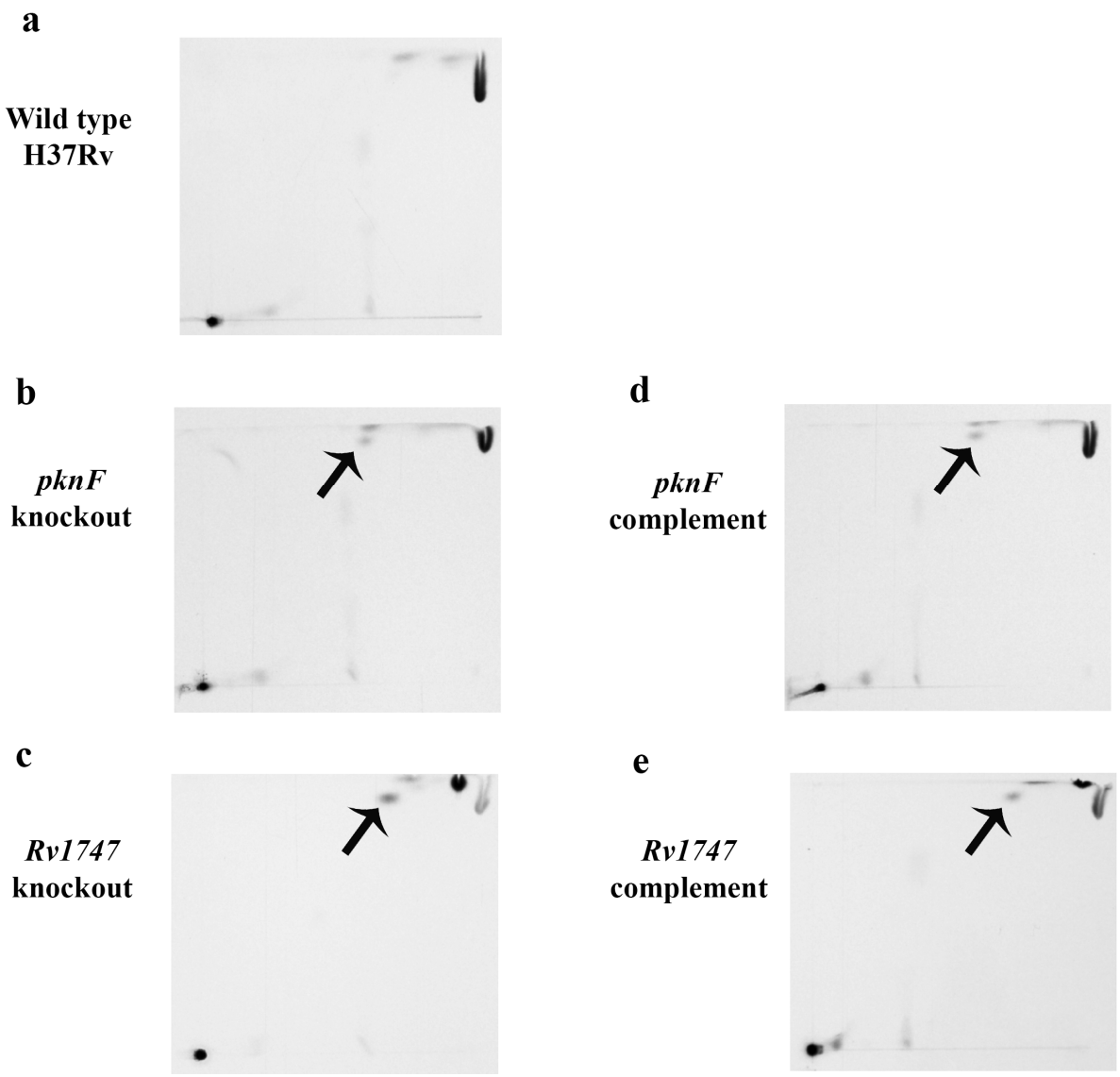


Figure 4.5 Two dimensional thin-layer chromatography of wild type H37Rv, *ΔpknF*, *ΔRv1747*, *pknF* complement and *Rv1747* complementing strains showing the polar lipids present in the cell pellets after separation in system D. Lipids were extracted from cell pellets as outlined in section 2.4.6. 20, 000 cpm of each sample was spotted onto a silica plate and then the lipids were separated in two different directions before being exposed to x-ray film. The black arrow highlights the accumulating lipid in the *ΔpknF* and *ΔRv1747* strains (figures 4.2b and 4.2c). Figures 4.2d and 4.2e show that the lipid is still present in the *pknF* and *Rv1747* complementing strains.

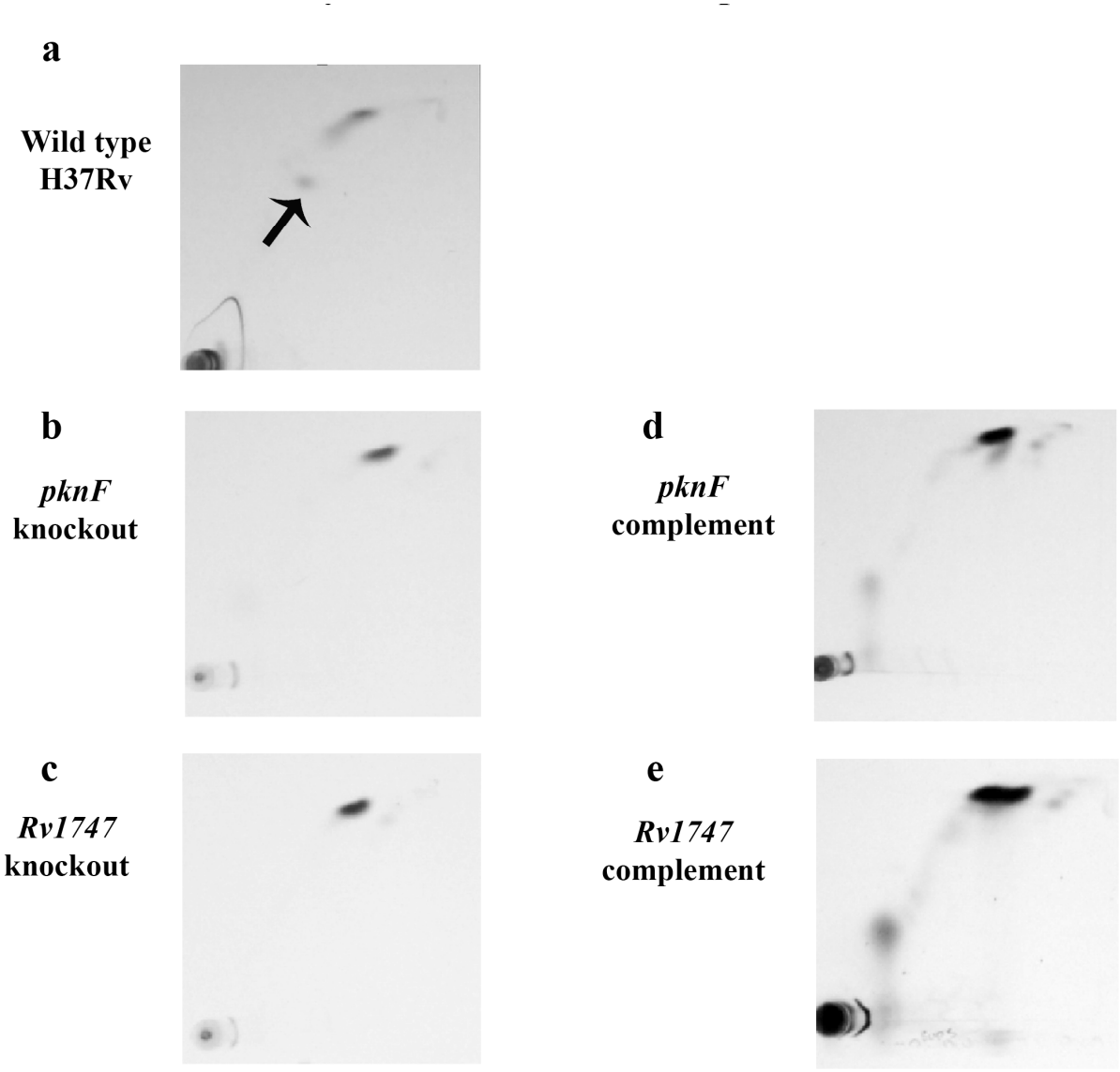


Figure 4.6 Two dimensional thin-layer chromatography of wild type H37Rv, *ΔpknF*, *ΔRv1747*, *pknF* complement and *Rv1747* complementing strains showing the polar lipids present in the culture supernatant of the *M. tuberculosis* strains after separation in system D. Lipids were extracted from *M. tuberculosis* supernatants as outlined in section 2.4.6. 20, 000 cpm of each sample was spotted onto a silica plate and then the lipids were separated in two different directions before being exposed to x-ray film. The black arrow highlights the accumulating lipid in the wild type H37Rv strain (figure 4.3a). Figures 4.3b to 4.3e show that the accumulating lipid is not present in either of the knockout or complementing strains.

4.3.1 Whole genome sequencing of wild type, *ΔRv1747* and *Rv1747* complementing strains of *M. tuberculosis*

Whole genome sequencing of *M. tuberculosis* strains was performed to assess whether there were any other mutations within the genome of the *Rv1747* deletion strain and its respective complementing strain, apart from the *Rv1747* gene deletion itself, which could account for the lipid phenotypes observed in the 2D TLC investigation. The wild type H37Rv parental strain was also sequenced so this could be compared with the deletion and complement strains. *M. tuberculosis* genomic DNA was prepared as outlined in section 2.4.2 and whole genome sequencing was performed by GATC Biotech. Single nucleotide polymorphisms (SNPs) were mapped for each of the three strains by comparing them to the original sequenced H37Rv strain (Cole *et al.*, 1998) with the help of Iñaki Comas (NIMR).

Figure 4.7 shows a Venn diagram summarising the distribution of SNPs found in and between the wild type, *ΔRv1747* and *Rv1747* complementing strains. All three strains shared 51 SNPs which differed from the H37Rv sequence (Cole *et al.*, 1998); it is known that the 1998 reference strain contains between 50 and 80 sequencing errors (Iñaki Comas, personal communication) and hence this is the reason the three strains share 51 SNPs. These SNPs were not important to this investigation as the focus of this analysis was on SNPs present only in individual strains and those shared between two strains. Table 4.7 shows each SNP identified after genome sequencing that was not shared between all three strains. Two types of SNP were identified in this study; a synonymous SNP, which does not change protein function as the nucleotide change does not result in an amino acid replacement, and a nonsynonymous SNP, which has the potential to alter protein function as the nucleotide change results in an amino acid substitution in the protein.

One SNP was found which was present in the wild type strain only. The SNP was identified as synonymous and was found to reside in Rv1200 which is a probable conserved integral membrane transport protein. There were no SNPs found only in the *ΔRv1747* strain but there were three SNPS present only in the *Rv1747* complement

which were all nonsynonymous. A SNP was found in *ctpA* (*Rv0092*) which is a probable cation transporter P-type ATPase, *Rv1640c* which is a lysyl tRNA synthetase enzyme that catalyses the reaction which joins a specific amino acid with a tRNA, and the third SNP was found in *Rv1747*.

Furthermore, two SNPs were found to be shared between the wild type and *ΔRv1747* strain. A synonymous SNP was identified in *drrA* (*Rv2936*) which is a probable daunorubicin ABC transporter and a nonsynonymous SNP was identified in *cpbK* (*Rv2070c*), probably involved in cobalamin biosynthesis. As both of these SNPs were shared between the wild type and deletion strain they could not have contributed to the lipid phenotypes observed in the TLC experiments.

Importantly, two SNPs were identified that were shared between the *Rv1747* deletion and complement strains. Both these SNPS were of interest because an alteration in protein function here could account for why the lipid phenotypes were seen and furthermore could offer an explanation as to why they could not be complemented. The first polymorphism was found in the intergenic region between two convergent genes, *Rv2813* and *Rv2814c*. Rv2813 is annotated as a conserved hypothetical protein and Rv2814c is a probable transposase. The second nonsynonymous SNP was found in Rv1748 at amino acid position 139 of the protein which resulted in a leucine residue being replaced with a serine. Rv1748 is a hypothetical protein which according to Tuberculist is possibly exported and has a single transmembrane helix between amino acid residues 23 to 45 (http://tuberculist.epfl.ch).

In summary, the results showed that there were two nonsynonymous SNPs shared between the *Rv1747* deletion and complement strains which may have offered an explanation as to why the lipid phenotypes were observed in the 2D TLC study and also provide an answer as to why they were not complemented.

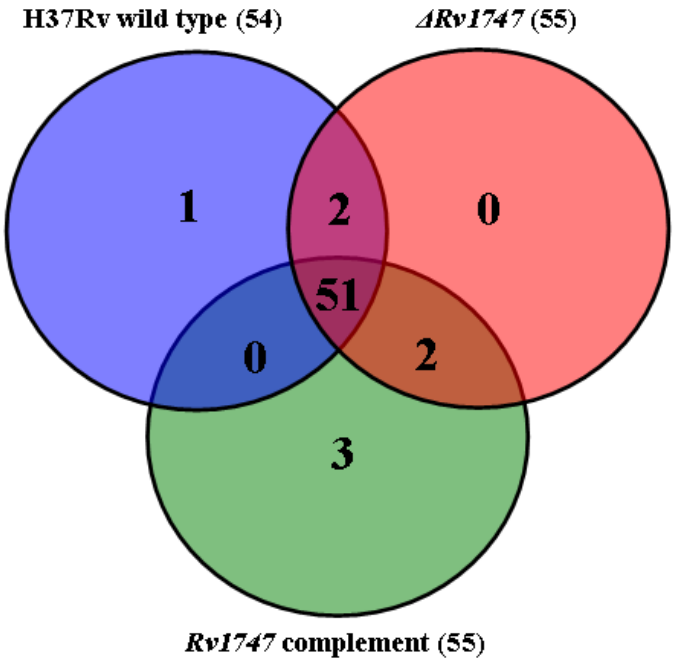


Figure 4.7 Summary of whole genome sequencing results. Whole genome sequencing of H37Rv wild type, *ΔRv1747* and *Rv1747* complement strains was performed by GATC Biotech as described in 2.4.2. The numbers in each section of the Venn diagram show how many single nucleotide polymorphisms were identified compared with the H37Rv reference strain sequenced in 1998 (Cole *et al.*, 1998). The blue circle represents the H37Rv wild type strain, the red circle is the *ΔRv1747* strain and the green circle is the *Rv1747* complement strain. The numbers in brackets after the name of each strain indicate the total number of SNPs found within that strain.

Strain	Gene	Gene function	SNP	Position on chromosome	Synonymous or nonsynonymous
SNPs in H37Rv wild type	Rv1200	Integral membrane transport protein	A-C	1344078	L379L: Synonymous
SNPs in *ΔRv1747*	-	-	-	-	-
SNPs in *Rv1747* complement	Rv0092/ctpA	Cation transporter ATPase	G-T	100706	V42F: Nonsynonymous
	Rv1640c/lysX	Lysyl-tRNA synthetase	G-A	1849437	P867S: Nonsynonymous
	Rv1747	ABC transporter	T-C	1974267	M213T: Nonsynonymous
SNPS shared by H37Rv wild type and *ΔRv1747*	Rv2070c/cobK	probably involved in cobalamin biosynthesis	C-G	2327933	R98P: Nonsynonymous
	Rv2936/drrA	Probable daunorubicin ABC transporter drrA	C-A	3272465	I84I : Synonymous
SNPs shared by *ΔRv1747* and *Rv1747* complement	Rv1748	Conserved hypothetical protein	T-C	1977015	L139S: Nonsynonymous
	Intergenic region between Rv2813 and Rv2814c	-	G-C	3119451	-
SNPs shared by H37Rv wild type and *Rv1747* complement	-	-	-	-	-

Table 4.7 Summary of SNPs identified from whole genome sequencing. Whole genome sequencing of H37Rv wild type, *ΔRv1747* and *Rv1747* complement strains was performed by GATC biotech as described in 2.4.2. Each SNP identified after genome sequencing that was not shared between all three strains is shown in the table.

4.4 Assessment of the lipoarabinomannan content of the mycobacterial cell wall

One hypothesis raised during this study was that perhaps the Rv1747 ABC transporter could transport lipoarabinomannan (LAM) to the outer membrane of the *M. tuberculosis* cell. LAM is an important immunomodulatory molecule of the cell surface of *M. tuberculosis* (like LPS is in Gram-negative bacteria) and it was proposed that to transport lipoglycans from the cytoplasm where they are synthesised to the outer layer of the cell envelope must require a yet unidentified ABC transporter (Pitarque *et al.*, 2008). This hypothesis was investigated using four antibodies raised against LAM and various capsular components of the *M. tuberculosis* cell wall in whole cell ELISAs. *M. tuberculosis* cell suspensions from wild type, *ΔRv1747* and *Rv1747* complement cultures were adsorbed to 96-well ELISA plates. After incubation with the primary antibodies an IgM secondary antibody peroxidise conjugate was added to each well. This antibody allowed for development with tetramethylbenzidine (TMB) substrate for horseradish peroxidise. Colour formation was monitored (wells turned blue in the presence of secondary antibody) and the reaction was stopped by adding H_2SO_4 to each well producing a yellow coloured product. Plates were then measured at 450 nm as described in section 2.4.14.

Unfortunately two antibodies did not work in the assay. These were Mab Baba which recognises capsular glucan and 55.92 1a1 which recognises ManLAMs capped with single mannosyl residues. Figure 4.8 shows the results from the whole cell ELISAs using the Mab F183-24 and Mab F30-5 antibodies. Mab F183-24 recognises ManLAM capped with three mannosyl residues and Mab F30-5 recognises the Ara6 structure present in LAM. Figure 4.9 shows a schematic structure of ManLAM highlighting the different proteins that each antibody recognised. Figure 4.8a shows the results of the ELISAs performed with Mab F183-24 and figure 4.8b plots the results of the ELISAs performed with Mab F30-5. The data in figures 4.8a and 4.8b show that there are no differences in the levels of ManLAM between the wild type, *ΔRv1747* and *Rv1747* complementing strains.

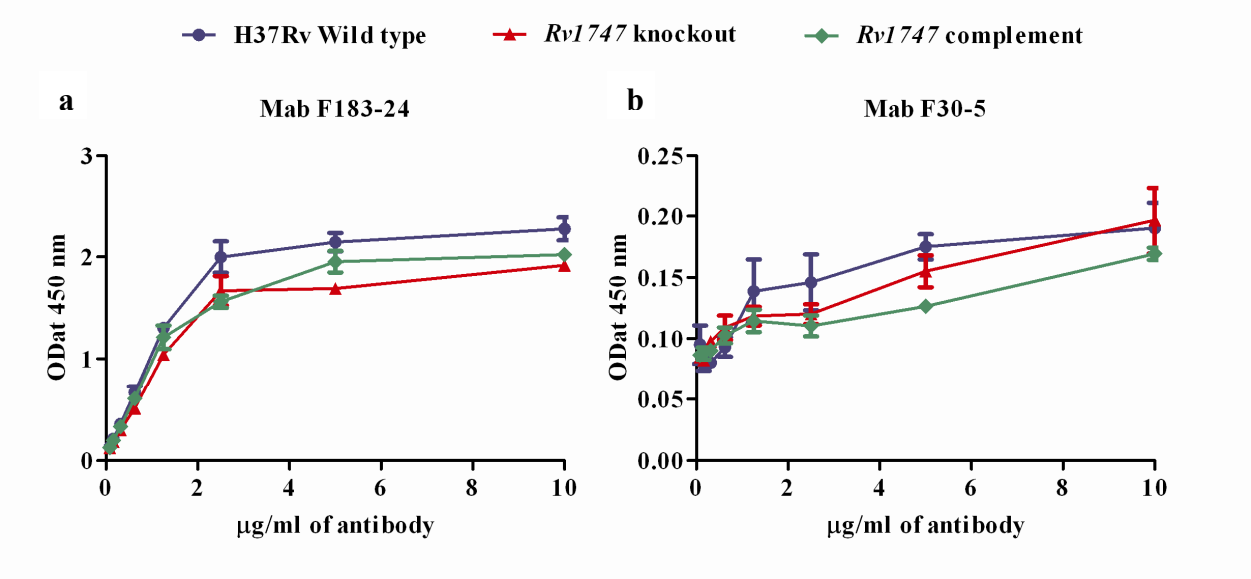


Figure 4.8 *M. tuberculosis* whole cell ELISAs comparing the levels of ManLAM in H37Rv wild type, *ΔRv1747* and *Rv1747* complement strains. *M. tuberculosis* whole cell ELISAs were performed as outlined in section 2.4.14. Data are plotted as the amount of primary antibody used in the ELISA in µg/ml versus the absorbance of the secondary peroxidise conjugated antibody at OD_{450nm} after development with TMB substrate. Figure a shows the data from the Mab F183-24 experiment and figure b shows the data from the Mab F30-5 experiment. Data are plotted as the average of three technical replicates and the error bars show the standard deviations.

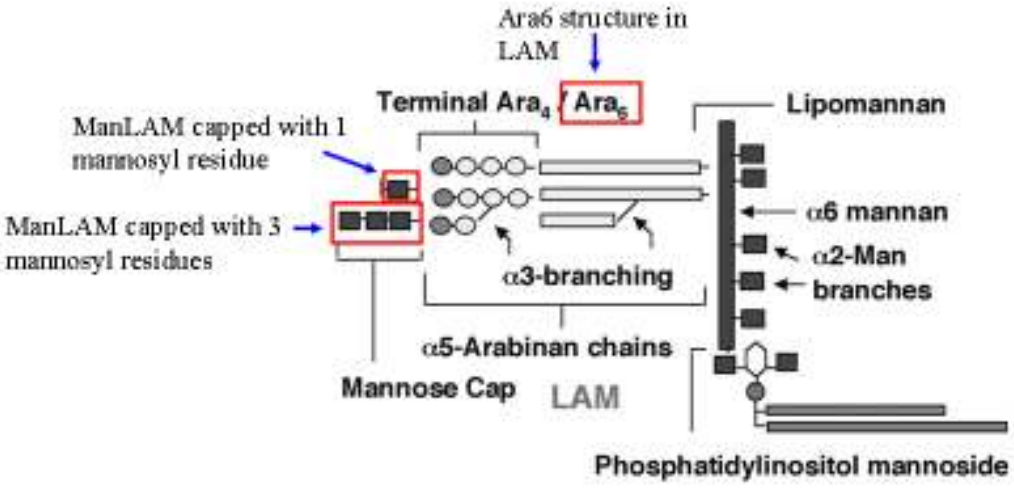


Figure 4.9 Schematic structure of mannose-capped LAM (ManLAM). The schematic shows a generalised figure of ManLAM showing the mannose cap, the mannose core and the phosphatidylinositol anchor. Highlighted by the three red boxes are the proteins which are recognized by the antibodies used in the LAM ELISAs. 55.92 1a1 antibody recognises ManLAMs capped with single mannosyl residues, Mab F183-24 recognises ManLAM capped with three mannosyl residues and Mab F30-5 recognises the Ara6 structure present in LAM. Figure adapted from Torrelles *et al* (Torrelles *et al.*, 2004)

4.5 Transmission electron microscopy of the *M. tuberculosis* cell wall

Transmission electron microscopy was performed to examine if there were any differences in the cell wall structure and composition between wild type, *ΔpknF*, *ΔRv1747* and the respective complementing strains. *M. tuberculosis* cultures were grown to exponential phase (OD_{600} 0.6-0.8) in Dubos medium before being resuspended in 2% glutaraldehyde / 2% paraformaldehyde in 0.1 M sodium cacodylate buffer (pH 7.2) for 18 hours in order to fix the cells. Cells were then prepared for TEM and sectioned by E. M. Hirst (NIMR). Images were then taken with a CCD camera at x 25,000 magnification. Ten images were taken per section and three sections were examined per strain.

Figures 4.10 and 4.11 show images taken of each *M. tuberculosis* strain. Both figures show a representative cross section of cells including both longitudinal and transverse sections. Figures 4.10 a, b and c show sections of wild type H37Rv cells; figures 4.10d, e and f show sections of *ΔpknF* cells and figures 4.9g, h and i show sections of *pknF* complement cells. In addition, sections c, f and i contain a zoomed image to highlight the cell wall structure. No discernable differences in cell wall structure were evident between the strains. Figures 4.11 a, b and c show sections of wild type H37Rv cells; d, e and f show sections of *ΔRv1747* cells; g, h and i show sections of *Rv1747* complement cells. Sections c, f and i contain a zoomed image to highlight the cell wall structure. Again, no discernable differences in cell wall structure were evident between the strains.

4.6 Assessment of the protein composition of *M. tuberculosis* culture supernatants

A hypothesis proposed during this study was that the Rv1747 transporter may export a protein required for the growth of the bacterium *in vivo*. Therefore, to assess this possibility the composition of culture supernatants were examined in the wild type, *ΔpknF*, *ΔRv1747* and respective complementing strains. *M. tuberculosis* cultures were

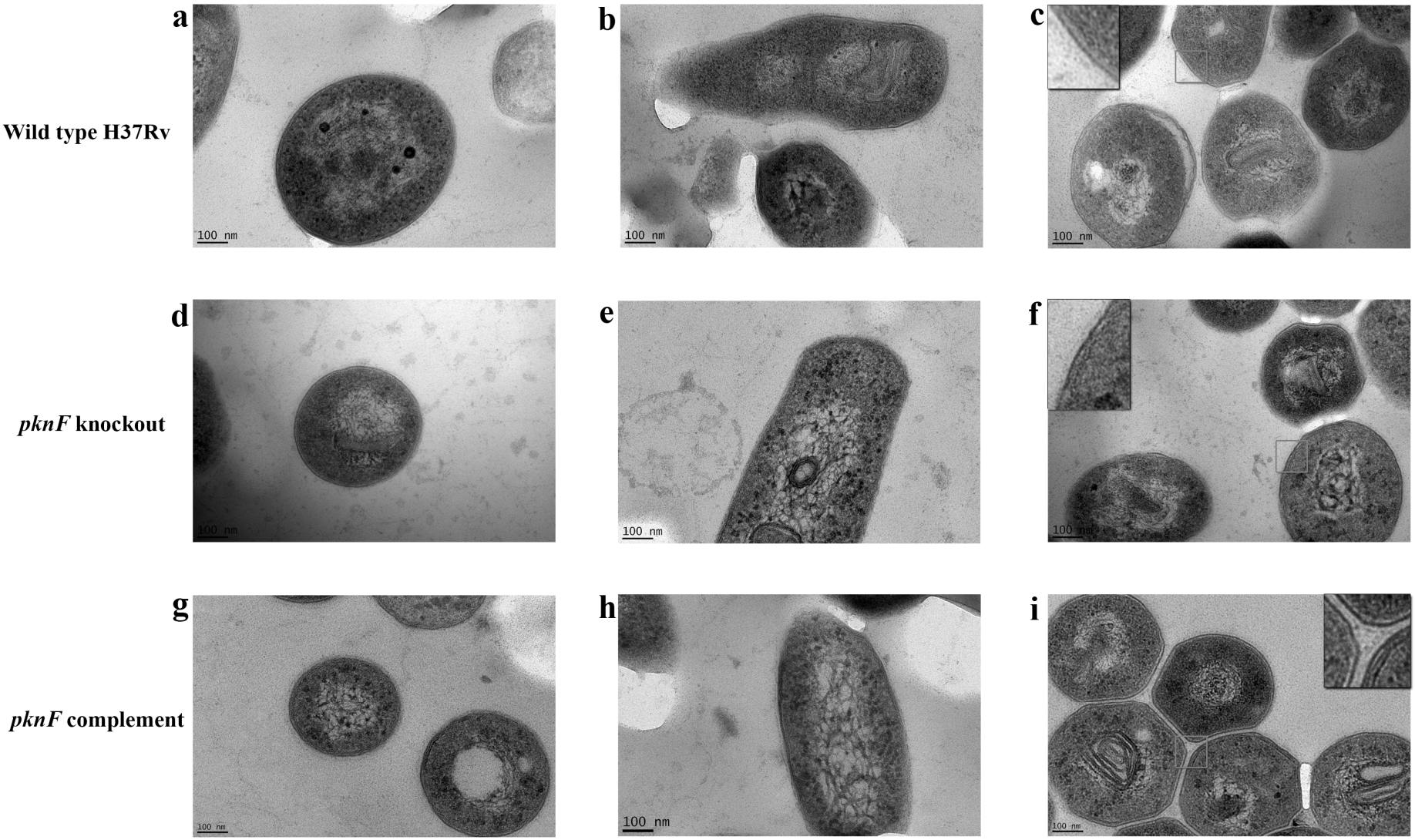


Figure 4.10 Transmission electron micrographs of *M. tuberculosis* comparing cell wall structure in wild type H37Rv, *ΔpknF* and *pknF* complement strains. Images were taken at x 25,000 magnification with a CCD camera. The scale bar on each image represents 100 nm. The figure shows a representative cross section of cells including both longitudinal and transverse sections. a, b and c show sections of wild type H37Rv cells; d, e and f show sections of *ΔpknF* cells; g, h and i show sections of *pknF* complement cells. Sections c, f and i contain a zoomed image to highlight the cell wall structure.

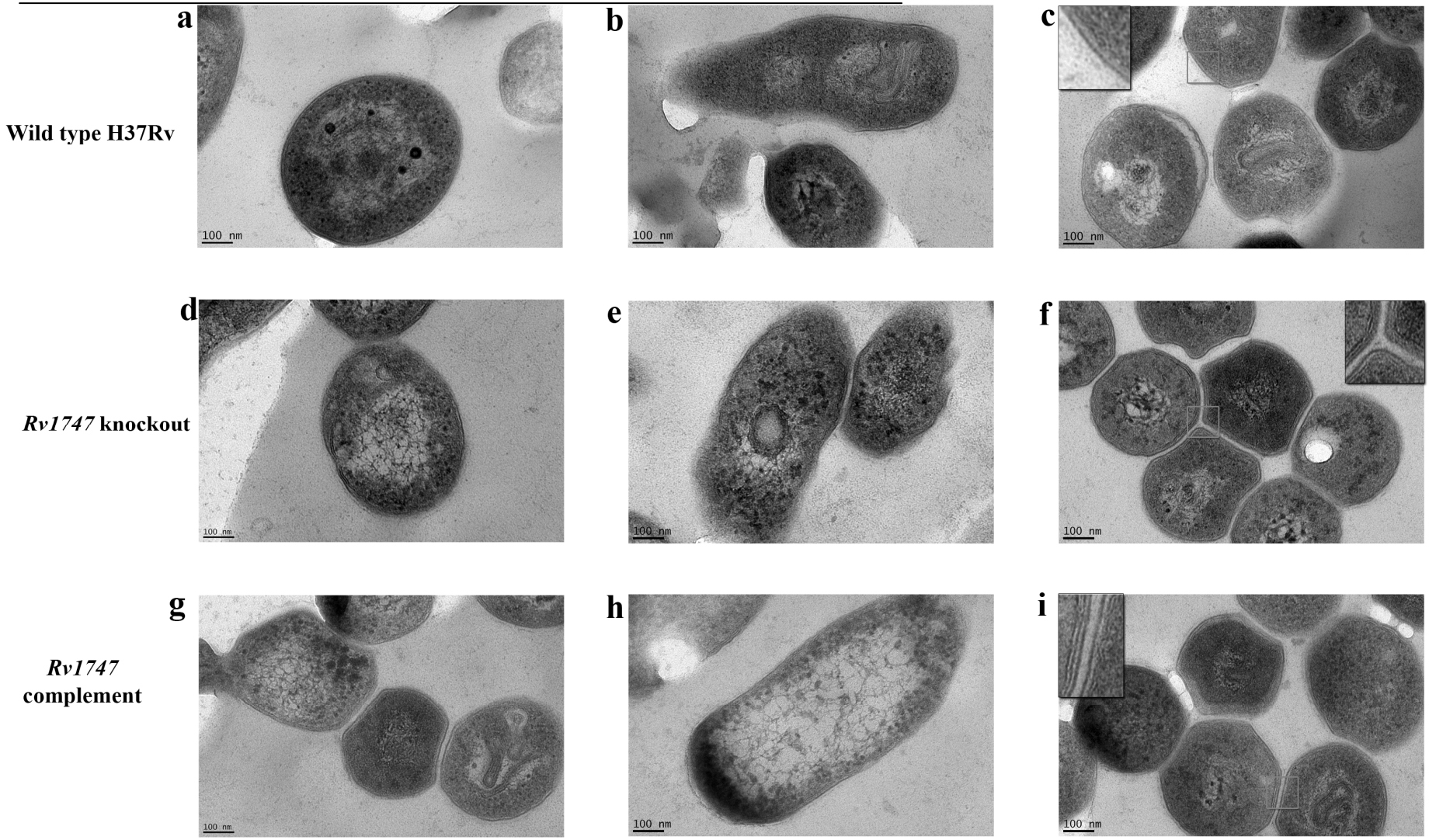


Figure 4.11 Transmission electron micrographs of *M. tuberculosis* comparing cell wall structure in wild type H37Rv, *ΔRv1747* and *Rv1747* complement strains. Images were taken at x 25,000 magnification with a CCD camera. The scale bar on each image represents 100 nm. The figure shows a representative cross section of cells including both longitudinal and transverse sections. Figures 4.10 a, b and c show sections of wild type H37Rv cells; d, e and f show sections of *ΔRv1747* cells; g, h and i show sections of *Rv1747* complement cells. Sections c, f and i contain a zoomed image to highlight the cell wall structure.

grown in Dubos medium to exponential phase (OD_{600} 0.6-0.8), cells were harvested and then the culture supernatants were prepared as described in section 2.4.4.1. Trichloracetic acid (TCA) was then used to precipitate the proteins from the *M. tuberculosis* culture supernatants; TCA treatment ensured each protein sample was as concentrated as possible to maximise the chances of observing any differences between the strains. Half of each TCA precipitated sample was then digested with trypsin; the enzyme is a serine protease which cleaves proteins at the carboxyl side of arginine or lysine. Samples were subjected to trypsin treatment to digest each protein into smaller fragments in case any differences in protein composition were hidden in the un-digested samples by a protein which made up a large proportion of the sample. Proteins were then separated by SDS-PAGE as described in 2.5.1.

Figure 4.12 shows the proteins present in the *M. tuberculosis* culture supernatants of all five strains after separation by SDS-PAGE. No significant differences in protein content can be observed between the strains after either treatment. The large abundance of a ~ 67 kDa protein observed in each lane is albumin, a component of the Dubos medium. Supernatants were also assessed for differences in protein content when *M. tuberculosis* cultures were at an OD_{600} of 0.3 and 1.0 (data not shown). There were no significant differences between the proteins present in the strains in either of these experiments.

4.7 Analysis of small molecules produced by *M. tuberculosis*

Nuclear magnetic resonance (NMR) spectroscopy experiments were performed to determine if mutation of *Rv1747* affected the small molecules or metabolites produced by *M. tuberculosis*. It was hypothesised that if the R1747 transporter exported a small molecule from the cell then we may have observed an absence of the molecule when examining the *ΔRv1747* strain compared with wild type and the complementing strain. Culture supernatants were harvested throughout growth (at an OD_{600} of 0.3, 0.6 1.0 and 2.0) and then analysed by NMR by Volker Beherends (Imperial College) as described in section 2.4.11. The small molecules and metabolites present in the supernatants of all three strains were the same throughout growth (data not shown).

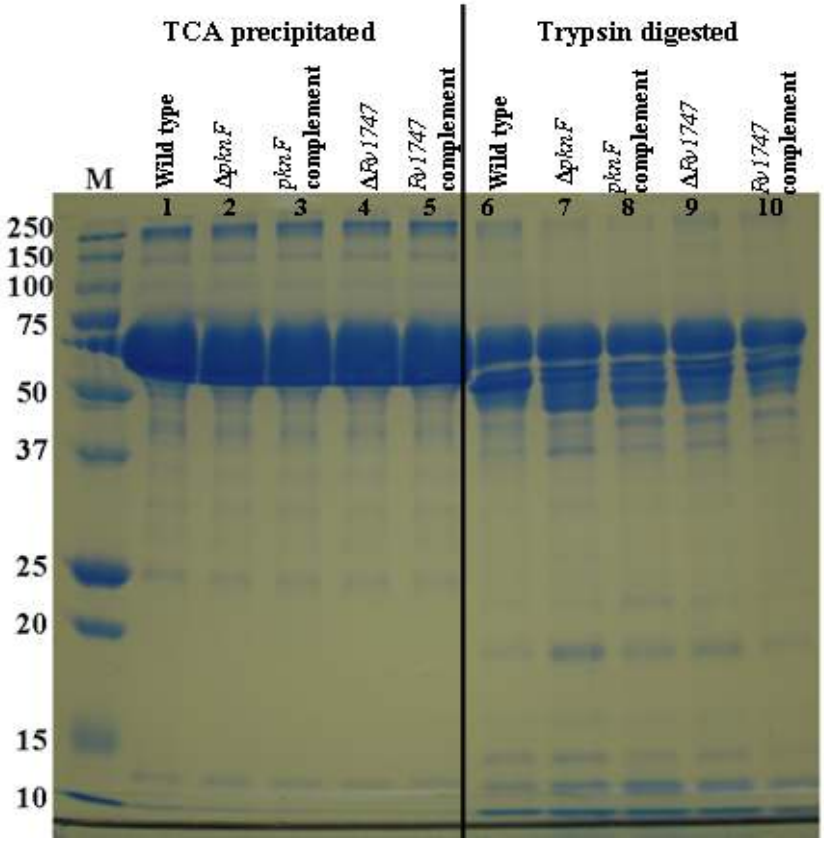


Figure 4.12 SDS-PAGE of proteins present in the supernatant of wild type *M. tuberculosis*, *ΔpknF*, *pknF* complement, *ΔRv1747*, and *Rv1747* complement strains. Protein extracts were prepared as outlined in section 2.4.4.1. Lanes 1 to 5 are trichloracetic acid precipitated samples and lanes 6 to 10 are trypsin digested samples. 20 µl of each sample was analysed by SDS-PAGE. Lanes 1 and 6 show supernatants from wild type H37Rv strains, lanes 2 and 7 show supernatants from the *ΔpknF* strain, lanes 3 and 8 show supernatants from the *pknF* complement strain, lanes 4 and 9 show supernatants from the *ΔRv1747* strain and lanes 5 and 10 show supernatants from the *Rv1747* complement strain. M is a protein marker and the relevant protein sizes are shown on the left hand side of the figure. No differences in protein content can be observed between the strains after either treatment.

4.8 Metabolic tests for *M. tuberculosis*: Phenotype MicroArrays™ (BiOLOG)

Phenotype MicroArrays™ were used to assess numerous cellular phenotypes of wild type, *ΔpknF*, *ΔRv1747* and the two complementing strains. The Phenotype MicroArrays used in this study tested the ability of *M. tuberculosis* to respire when exposed to a range of substrates which included carbon, nitrogen, phosphorus and sulphur sources, various nutrient supplements, peptide nitrogen sources and osmolytes and also tested the ability of the cells to respire in a range of pH conditions. It was hypothesised that differences found in substrate usage may yield clues as to the function of the two proteins and may help to identify the substrate for the Rv1747 ABC transporter.

M. tuberculosis strains were prepared according to section 2.4.10 and then added to each well of the arrays and incubated for seven days at 37 ^{o}C. In the presence of respiring cells a purple coloured compound formed from the reduction of a tetrazolium dye (Biolog's patented redox chemistry) which was added to the inocula. On day seven the A570 of each well was measured.

Pair-wise comparisons were made between the wild type strain and both the mutant strains and the wild type strain versus both the complement strains. The absorbance of each plate was normalised to the average absorbance across the entire plate to correct for any differences in growth between the plates. The background absorbance from the negative control well (cells but no substrate) was then subtracted. Finally, fold changes were calculated for substrates where at least one of the strains had an absorbance higher than the negative control well. Results showed that there were no reproducible differences in substrate usage found between the wild type and mutant strains (data not shown).

4.9 Discussion

The aim of this study was to identify the substrate transported by Rv1747. Previous work in the laboratory highlighted the important function of this protein in macrophage

and mouse models of infection thus warranting further analysis in this investigation (Curry *et al.*, 2005).

4.9.1 Transcriptional Microarrays and qRT-PCR

Microarray data showed that there were significant changes in the transcriptional profiles between wild type strains and both the mutant strains. Interestingly, the gene most up-regulated in both of the mutant strains was *iniB*. Furthermore, the *iniA* gene was also up-regulated in both the mutants whilst the *iniC* gene was significantly up-regulated in the *ΔRv1747* strain only. The *iniBAC* genes were first identified as three isoniazid-induced genes in a screen in 1998 which assessed gene expression changes in *M. tuberculosis* in response to isoniazid treatment (Alland *et al.*, 1998). Furthermore, this study showed that the *iniA* gene was also induced by ethambutol which is another *M. tuberculosis* therapeutic that also targets the cell wall but has a different mechanism of action compared to isoniazid. In a later study using *M. bovis* BCG the promoter of the *iniBAC* operon was shown to be specifically induced by a broad range of inhibitors to cell wall biosynthesis including antibiotics that inhibited the synthesis of peptidoglycan, arabinogalactans, mycolic acids and fatty acids (Alland *et al.*, 2000). Using a luciferase reporter assay they demonstrated that the *iniBAC* promoter was induced 10 to 30 fold following a 24 or 48 hour treatment with inhibitory concentrations of a range of cell wall targeted drugs, including isoniazid, ethambutol and 5-chloropyrazinamide. In addition, a 3 to 5 fold induction was observed after treatment with ampicillin which targets peptidoglycan. They also showed that this induction was only possible when treating actively growing cells. The same study then showed the *iniBAC* promoter was not induced or repressed by other stress conditions including hydrogen peroxide, heat shock, acid stress, kanamycin, ciprofloxacin and rifampin, all of which do not directly target cell wall biosynthesis but are toxic to the cell via other mechanisms. Furthermore the luciferase reporter activity did not increase when cells were treated with lysozyme showing that cell lysis did not induce the promoter (Alland *et al.*, 2000). This suggested that the increased *iniBAC* promoter activity was a specific response to cell wall biosynthesis stress or damage. In 2005, a study was published which showed that *iniA* was essential for the activity of an efflux pump which conferred resistance to isoniazid

and ethambutol (Colangeli *et al.*, 2005). In this study *M. tuberculosis iniA* was over expressed in *M. bovis* BCG and this strain was shown to survive longer upon treatment with inhibitory concentrations of isoniazid and ethambutol compared with the control. Conversely, an *M. tuberculosis iniA* deletion strain had increased susceptibility to treatment with isoniazid. Using these two strains Colangeli and colleagues then showed the overexpression strain had a decreased accumulation of ethidium bromide compared with the control whereas the deletion strain accumulated ethidium bromide. What is more, these effects could be reversed using an efflux pump inhibitor. The pump inhibitor also reversed the resistance to isoniazid observed in the overexpression strain. They concluded that *iniA* is a pump component but does not directly transport isoniazid from the cell (Colangeli *et al.*, 2005). All these findings would be compatible with the Rv1747 transporter exporting a component of the cell wall necessary for full virulence of the bacillus.

Many other significant changes in gene expression were found in the microarray study. For the *ΔRv1747* study these most notably included the expression of *Rv1747* which was 29 fold down-regulated in the *ΔRv1747* strain compared to wild type. In the *Rv1747* complement strain expression of *Rv1747* was 1.3 fold up-regulated compared to wild type confirming the restoration of gene expression. This result was important as it confirmed (along with the qRT-PCR results) that there was no expression of *Rv1747* in the null strain and demonstrated that the abundance of transcript returned to wild type levels in the complement. Therefore, using the *Rv1747* null strain with no presence of transcript would allow for reliable experiments to be performed to try and identify the transported substrate. Furthermore, the fact that the levels of transcript in the complemented strain were very similar to wild type levels meant that the complement could be used in experiments to determine if any differences seen between the wild type and deletion strains were due to the deletion of *Rv1747* or were the result of another genetic difference between the strains.

Other significant changes in gene expression in the *ΔRv1747* microarray study included *Rv0047c*, *fadE23*, *fdxA*, *Rv0822c*, *gyrB*, *PE_PGRS41*, *ethA* and *ino1*, all of which were in the top ten list of up-regulated genes in the study. All these genes were found to be

up-regulated two fold or more. A two fold cut-off level was applied to the data because any changes smaller than two fold were deemed to be probably not significant in terms of gene function. EthA functions to activate the pro-drugs ethionamide, thiacetazone and isoxyl which all use different mechanisms to inhibit mycolic acid synthesis (Dover *et al.*, 2007). Mycolic acids are another important constituent of the *M. tuberculosis* cell wall. Ino1 is involved in the phosphatidyl-myo-inositol (PI) biosynthetic pathway and was up-regulated 2.0 fold in the *Rv1747* deletion strain. This phospholipid is also a component of the cell envelope and can be seen in the 2D TLC results (figure 4.4 system E). However, the profile of this lipid did not change between the wild type and mutant strains (figure 4.4). Consequently, the fact that *ethA* and *ino1* were also both up-regulated in the *ΔRv1747* strain, like the *iniBAC* operon, added further weight to the hypothesis that Rv1747 functioned to transport an important component of the *M. tuberculosis* cell wall; furthermore, up-regulation of all these genes may be acting as a compensatory mechanism for the loss of the function of Rv1747. The list of the top ten down-regulated genes included two genes annotated as being conserved integral membrane proteins, namely *Rv1999c* and *Rv2265*, and one gene, *Rv1382*, annotated as a probable export or membrane protein. This is perhaps interesting as Rv1747 is also a membrane protein with an export function; however, little is known about the function of these genes. *Rv1382* was predicted to be an essential gene in the Sassetti transposon mutagenesis screen (Sassetti *et al.*, 2003).

Microarray data also showed that *pknF* was 2.0 fold up-regulated in the *ΔRv1747* strain and was still 2.0 fold upregulated in the *Rv1747* complement. The level of up-regulation seen in the *Rv1747* complement is most likely because the *Rv1747* complementing plasmid included a copy of the *pknF* gene to ensure that all the upstream elements required for normal expression of *Rv1747* were there, and to make certain the promoter was also present. Having an extra copy of *pknF* integrated into the *M. tuberculosis* genome in the *Rv1747* complement strain would mean that it would be likely that there would be twice the amount of *pknF* transcript present, hence when performing the microarrays and qRT-PCR this fact would account for the elevated levels of *pknF* transcript. In the qPT-PCR analysis however, the levels of *pknF* transcript were not increased in the *ΔRv1747* strain, unlike the microarray data which showed the transcript

was 2.0 fold up-regulated. This highlights why microarrays should only be used as a screening tool to generate hypotheses and must always be backed up by another method to confirm gene expression changes.

In the *pknF* microarray experiment there were fewer genes whose expression level altered significantly upon deletion of *pknF*. This was particularly striking when a two fold level of change cut-off was applied to the data and it was discovered that there were only 12 genes that were differentially regulated between the wild type and mutant strain compared with the 40 genes which passed the cut-off in the *Rv1747* experiment. In addition all 12 genes were found to be down-regulated. Due to the low number of differentially regulated genes passing a two fold cut-off the stringency was reduced to 1.5 fold. With a 1.5 fold cut-off the number of genes found to be differentially regulated rose to 72. Yet again 72 differentially regulated genes is a very small number passing a 1.5 fold cut-off compared with the 483 genes which passed in the *ΔRv1747* microarray analysis. One possible explanation for this finding could be because of crosstalk between the STPKs. Thus, PknF is one of 11 serine-threonine protein kinases present in *M. tuberculosis*. It has been shown that multiple kinases can phosphorylate the same substrate *in vitro* and that one kinase can phosphorylate multiple substrates. For example, in a 2005 study PknB, PknD, PknE and PknF were all shown to be able to phosphorylate an Rv1747 expression construct (Grundner *et al.*, 2005). Furthermore, PknF has been shown to phosphorylate another FHA domain containing protein, Rv0020c (Grundner *et al.*, 2005). Additionally, Molle *et al* performed *in vitro* assays to demonstrate that PknF phosphorylates KasA, KasB (Molle *et al.*, 2006a) and the GroEL1 chaperone (Canova *et al.*, 2009). The GroEL1 chaperone was also found to be phosphorylated by PknA, B, D, E, H and L although to a lesser extent than PknF. Furthermore, in 2009 the same group showed PknF phosphorylates mtFabH to negatively regulate the FASII cycle involved in mycolic acid synthesis (Veyron-Churlet *et al.*, 2009). Although these studies were performed *in vitro*, which is an extremely unnatural environment for the proteins to be in and does not reflect the intracellular conditions faced by the bacterium including spatial and temporal separation, these data suggest there may be some level of redundancy between the kinases. Therefore deleting one kinase from the genome may mean that others can compensate for the effects of the

deletion. This may explain why we see fewer genes differentially expressed in the *pknF* deletion compared with the *Rv1747* deletion strain. However, what was interesting from this study was the observation that *iniB* and *iniA* were also up-regulated in the *ΔpknF* strain and well as the *ΔRv1747* strain.

Although the levels of fold change were lower in the *ΔpknF* data set, genes present in the *ΔpknF* list that were also differentially regulated in the *Rv1747* null strain included *fdxA* (1.8 fold up-regulated) and *ethA* (1.7 fold up-regulated). *fdxA* is a probable ferredoxin involved in electron transfer and is involved in the response to hypoxia in *M. tuberculosis*. The *fdxA* gene is induced 22 fold by hypoxia suggesting it is involved in survival in hypoxic conditions such as those faced by *M. tuberculosis* within the host in granulomas (Sherman *et al.*, 2001). Furthermore, seven genes in the top ten list of genes most highly down-regulated upon *pknF* deletion were putative transposases. Transposases are enzymes which catalyse the movement of a transposon to another part of the genome and there are more than 30 transposable elements within the genome of *M. tuberculosis* (Guilhot *et al.*, 1999).

In the microarrays comparing expression levels between wild type and *ΔpknF* strains the *pknF* gene itself did not pass the filtering in the analysis steps. This could have been due to anomalous hybridisation of cDNA to the *pknF* spot on the microarray slides. This again highlights why using qRT-PCR to substantiate microarray results is important. qRT-PCR results confirmed that expression of *pknF* in the null mutant was undetectable and the normalised transcription level was restored to almost wild type levels in the complementing strain. Additionally, in all qRT-PCR data the normalised level of transcription of *Rv1747* was higher than that for *pknF*. This is unusual because generally the transcript abundance decreases from the first gene to the last gene of an operon as is shown in the qRT-PCR results for the *iniBAC* operon. This highlights that the mechanism of regulation of *Rv1747* might be more complex than first thought. However, previous work demonstrated that it was unlikely that there is a promoter present for *Rv1747* (section 3.1 and 3.2) and in addition a Solexa sequencing trace of the *M. tuberculosis* transcriptome showed that the mRNA transcript abundance does not increase between *pknF* and *Rv1747* (Kristine Arnvig, personal communication).

4.9.2 The search for the Rv1747 substrate

Results from the microarray studies led to the hypothesis that the Rv1747 transporter could export a component of the *M. tuberculosis* cell wall necessary for growth of the bacterium *in vivo*. The *M. tuberculosis* cell wall is highly complex and is rich in lipids which make up to 60 percent of the dry weight of the bacteria (Kremer & Besra, 2005) and are critical for pathogenicity. The extreme hydrophobicity of the cell wall accounts for its low permeability and resistance to many drugs and chemotherapeutics. Therefore the first group of experiments in this study investigated whether Rv1747 exported a lipid from the cell, then explored whether there were any changes in lipoarabinomannan content between the strains and also examined the cells by transmission electron microscopy to determine whether there were any structural differences between the wild type and mutant strains.

Two dimensional thin-layer chromatography showed that there was a lipid accumulating in the cell pellets of the mutant strains in system D. Also, in the culture supernatants there was a lipid present in the wild type strain, in the same system, that was absent in the mutant strains. However, these reciprocal differences could not be complemented. It is possible that the lipid accumulation observed in both the *ΔpknF* and *ΔRv1747* cell pellets was because of an inactive (in the case of the *ΔpknF* strain) or deleted ABC transporter (in the case of the *ΔRv1747* strain). This theory assumes that there is a positive mechanism of regulation between the kinase and the transporter, i.e. phosphorylation of Rv1747 by PknF activates the transporter activity of Rv1747. The fact that microarray data showed similar patterns of gene expression changes is compatible with this theory. It could then be hypothesised that although the levels of transcription were restored to wild type levels or above in both the *Rv1747* and *pknF* complemented strains perhaps this did not translate into functional protein and therefore the lipid phenotype was not complemented. However, although an *in vivo* experiment, the *Rv1747* deletion could be fully complemented in a mouse intravenous injection model of tuberculosis (Curry *et al.*, 2005). An alternative explanation therefore was that the lipid differences observed were the result of another genetic difference between the wild type, mutant and complementing strains. To test this hypothesis the wild type,

ΔRv1747, and respective complement strain were sequenced by whole genome sequencing.

Whole genome sequencing of *M. tuberculosis* strains was performed to assess whether there were any other mutations within the genome of the *Rv1747* deletion strain and its respective complementing strain, apart from the *Rv1747* gene deletion itself, which could account for the lipid phenotypes observed in the 2D TLC investigation. The wild type H37Rv parental strain was also sequenced so this could be compared with the deletion and complemented strains. Additionally, there was also the possibility that the lipid phenotypes seen in the TLC investigations could have been caused by a mutation in the wild type strain that occurred after the construction of the both the deletion strains.

One SNP was found which was present in the wild type strain only. A polymorphism in the wild type strain could have explained why the lipid phenotypes seen in the other four strains were not seen in this strain. However, the SNP was identified as synonymous and was therefore not of interest to this study. There were no SNPs found only in the *ΔRv1747* strain but there were three SNPS present only in the *Rv1747* complement strain which were all nonsynonymous. A SNP was found in *ctpA* (*Rv0092*) which is a probable cation transporter P-type ATPase, *Rv1640c* which is a lysyl tRNA synthetase enzyme that catalyses the reaction which joins a specific amino acid with a tRNA to form an aminoacyl tRNA, and the third SNP was found in *Rv1747*. The first two SNPs described would not have affected the complementation or function of Rv1747 and furthermore, they would not have caused the lipid phenotype seen in both the mutant and complement strains as they were not present in the mutant strain. However, the *Rv1747* SNP in the complementing strain could have potentially meant that the function of Rv1747 was not complemented if indeed the phenotype which was seen was due to the deletion of *Rv1747*. The SNP was at amino acid position 213 and replaced a methionine residue with a threonine. Position 213 is situated in the Rv1747 protein between the FHA-2 domain and the nucleotide binding domain; this SNP is not in a functional domain of the protein. The mutation most probably occurred during the construction of the *Rv1747* complementing plasmid. It could be argued that this SNP was the reason that the lipid phenotype observed in the *Rv1747* deletion strain was not complemented. However, as

the *Rv1747* complement is able to rescue the phenotype seen in macrophage and murine infection experiments and coupled with the fact that the mutation is not in one of the annotated domains, it further substantiates the proposal that this SNP did not affect the function of the ABC transporter. Furthermore the SNP would also not explain why the same lipid phenotype was observed in the *ΔpknF* and *pknF* complement strains.

Two SNPs were found to be shared between the wild type and *ΔRv1747* strain. A synonymous SNP was identified in *drrA* (*Rv2936*) which is a probable daunorubicin ABC transporter and a nonsynonymous SNP was identified in *cpbK* (*Rv2070c*), probably involved in cobalamin biosynthesis. As both of these SNPs were shared between the wild type and deletion strain they could not have contributed to the lipid phenotypes observed in the TLC experiments.

Importantly, two SNPs were identified that were shared between the *Rv1747* deletion and complemented strains. Both these SNPS were of interest because an alteration in protein function in these strains could account for why the lipid phenotypes were seen and furthermore could offer an explanation as to why they could not be complemented. The first polymorphism was found in the intergenic region between two convergent genes, *Rv2813* and *Rv2814c*. Rv2813 is annotated as a conserved hypothetical protein and Rv2814c is a probable transposase. As the two genes are transcribed convergently it is not possible for a promoter or regulatory region for either gene to be present in the intergenic region; therefore the SNP could not influence the regulation of either of the genes in this way. However, although the SNP was found in an intergenic region it was possible that there may have been a small RNA present in this region, which was not annotated, which could have had regulatory effects on various genes if the SNP was nonsynonymous and affected its function. Upon further examination of a Solexa sequencing trace of the *M. tuberculosis* transcriptome in exponential phase, the mRNA transcript level of the intergenic region was found to be minimal providing further evidence that that there is no gene present in this region (Kristine Arnvig, personal communication). Therefore this SNP is not likely to have caused the phenotypes seen in the TLC study.

The second nonsynonymous SNP was found in Rv1748 at amino acid position 139 of the protein which resulted in a leucine residue being replaced with a serine. Rv1748 is a hypothetical protein, which is 243 amino acids in size, and according to Tuberculist is possibly exported and has a single transmembrane helix between amino acid residues 23 and 45 (http://tuberculist.epfl.ch). This mutation was most likely introduced into the *ΔRv1747* targeting construct and therefore when homologous recombination occurred to produce the *Rv1747* deletion the *Rv1748* SNP was introduced. *Rv1748* is 732 bp in length and the intergenic region between *Rv1747* and *Rv1748* is 337 bp. The *Rv1747* targeting construct included ~ 1.5 kb of homologous DNA on either side of the deletion meaning that the whole *Rv1748* gene was included in the *ΔRv1747* targeting construct. It can be hypothesised that if Rv1748 is involved in lipid metabolism then perhaps this could offer an explanation for the phenotypes observed. To test this hypothesis an *Rv1747* complementing plasmid could be designed which included *pknF*, *Rv1747* and *Rv1748*. This plasmid could be transformed into the *ΔRv1747* strain and 2D thin-layer chromatography could be repeated; one would expect that if Rv1748 was responsible for the phenotypes then they would be restored in this complementing strain.

However, the whole genome sequencing results did not immediately offer an explanation as to why the lipid phenotypes were also observed in the *ΔpknF* and *pknF* complement strains (figures 4.4 to 4.6). It was therefore assessed whether the Rv1748 SNP could account for all the lipid phenotypes observed. It was discovered in this study that *pknF* is only co-transcribed with *Rv1747* and does not also form an operon with *Rv1748* (figure 3.1). Therefore, a deletion of *pknF* would not affect the regulation of transcription of the *Rv1748* gene. Furthermore, the *pknF* deletion was constructed as an in-frame deletion and therefore there should have been no downstream effects on *Rv1747* transcription or effects on subsequent genes. Indeed, this was demonstrated by qRT-PCR which showed that the levels of transcription of *Rv1747* in the *ΔpknF* strain were the same as in the wild type and *pknF* complement strain (figure 4.3b). In addition the *pknF* deletion targeting construct only included the first 1654 bp of *Rv1747* and therefore there was no part of the *Rv1748* gene within the deletion plasmid. Consequently it can be assumed that there was not an *Rv1748* mutation introduced into the *pknF* deletion strain when the targeting vector was introduced into the genome.

However, it was not unreasonable at this stage to hypothesise that a nonsynonymous mutation in *Rv1748* could have occurred within the *pknF* deletion strain and may be the explanation as to the observed lipid phenotypes. Furthermore, if correct, the phenotypes would not have been restored by replacement of the *pknF* gene. As previously discussed, because Rv1748 has no categorised function it could be that this protein was involved in the metabolism of a lipid that was seen accumulating in the two deletion and two complementing strains. A nonsynonymous mutation in Rv1748 could prevent correct metabolism and therefore it could have meant that the lipid was not processed correctly in order for it to be exported from the *M. tuberculosis* cell. A SMART domain search revealed that Rv1748 has a single transmembrane domain and no other domains of homology demonstrating that Rv1748 itself cannot transport substrates due to a lack of a nucleotide binding domain and several transmembrane domains. Furthermore, a BLAST search revealed that Rv1748 has no strong homology with any other bacterial proteins whose function has been characterised hence no clues could be found as to the function of this protein. Additionally Rv1748 is predicted to be non-essential in the Sassetti mutagenesis screen (Sassetti *et al.*, 2003). Therefore the predicted function of the protein remains unknown. To investigate the hypothesis that a mutation in Rv1748 was the cause of the lipid phenotypes, the *Rv1748* gene plus its promoter region was sequenced in the *ΔpknF* strain (data not shown). Upon examination of the sequencing results they revealed that there were no mutations within the *Rv1748* gene or the 373 bp intergenic region between *Rv1747* and *Rv1748* thus suggesting that the lipid phenotype observed in the *ΔpknF* and *ΔRv1747* deletion strains were not due to rendering Rv1748 non-functional.

Assessment of the *ΔpknF* and *pknF* complemented strains in terms of PknF protein function is extremely difficult as no phenotype of this mutant has been discovered or characterised. Therefore, one cannot be certain whether the function of this kinase is restored in the *pknF* complementing strain as there is no measurable readout of phenotype in the mutant; however, at the transcriptional level the relative abundance of *pknF* transcript was returned to wild type levels in the *pknF* complemented strain (figure 4.3a) and the most highly upregulated gene in the deletion strain, *iniB*, was restored to wild type levels in the complementing strain (figure 4.3c). Both of these results suggest

that the function of the kinase is restored in the *pknF* complementing strain. Obviously one has to bear in mind that levels of transcription do not always correlate to levels of protein expression and furthermore cannot indicate levels of protein function.

Another hypothesis which could explain the phenotypes observed in the lipid study is that there could be a nonsynonymous mutation that is shared between all the deletion and complementing strains in one or more of the PE/PPE/PGRS genes. The PE and PPE gene families are highly repetitive proteins containing many glycine and alanine residues and constitute approximately ten percent of the genome in H37Rv (Marri *et al.*, 2006). *M. tuberculosis* has approximately 60 genes belonging to the PPE family and another 100 genes which are part of the PE or PE-PGRS family (Marri *et al.*, 2006). Members of the PE/PPE family are known to be cell wall associated proteins which can have effects on the immune response to infection (Voskuil *et al.*, 2004). Due to their inherent repetitiveness throughout the genome these genes have to be removed from the re-sequencing data set prior to analysis. Therefore it is also possible that there may be an unidentified mutation in one of these genes in both the deletion and complementing strains that resulted in the lipid phenotypes observed.

However, it still remains an alternative hypothesis that the lipid phenotype *was indeed* a consequence of *Rv1747* deletion (or Rv1747 inactivation in the case of the *ΔpknF* strain) if PknF positively regulates Rv1747, and the reason the lipid accumulation was still present in the *Rv1747* complementing strain could be due to incomplete complementation. Moreover, the *Rv1747* complementing plasmid included an extra copy of *pknF* (see figure 1.12) which may have affected complementation, and secondly, it has recently been discovered that *Rv1747* has a ~150 bp 3' untranslated region which may play a role in transcript stability and this was not included in the *Rv1747* complementing plasmid (K. Arnvig, Personal Communication) (figure 1.12).

In summary, **it cannot be completely eliminated that the lipid phenotypes observed in the 2D-TLC investigation were the result of incomplete complementation of the *Rv1747* and *pknF* deletion strains, thus implicating this system in lipid synthesis or transport. A viable alternative explanation however, is that there is a mutation in**

one or more of the *PE/PPE* gene family in both the *ΔpknF* and *ΔRv1747* mutant strains.

Lipoarabinomannan (LAM) is an important immunomodulatory molecule of the cell surface of *M. tuberculosis*. A study by Pitarque *et al* in 2008 labelled *M. bovis* BCG cell surface lipoglycans with biotin and found that a group of them, including lipoarabinomannan, were exposed at the cell surface and thus were presumably inserted into the outer most membrane of the bacterium (Pitarque *et al.*, 2008). They then proposed that to transport the lipoglycans from the cytoplasm where they are synthesised to the outer layer of the cell envelope must require a yet unidentified ABC transporter. In the 2008 study Pitarque and colleagues tested one possible candidate, a putative polysaccharide ABC transporter encoded by *Rv3781* and *Rv3783*. The *Rv3783* ortholog in *M. smegmatis* was deleted by homologous recombination; however, this strain secreted similar amounts of LAM, glucan, arabinomannan and mannan. Furthermore, in the mutant, lipoglycans were as effectively tagged by biotin as the wild type parental strain leaving the identity of the transporter unknown (Pitarque *et al.*, 2008). It was therefore hypothesised that Rv1747 could function to transport LAM to the outer layer of the *M. tuberculosis* cell envelope. In this present study ELISAs were performed to investigate this hypothesis. However, results showed that there were no differences in the levels of LAM and its components between the wild type, *ΔRv1747* and *Rv1747* complementing strains conclusively showing that **Rv1747 does not export LAM to the mycobacterial outer cell surface**. Unfortunately, two of the antibodies did not work, presumably because the antibody protein became denatured in transit and thus I was unable to detect the components of the *M. tuberculosis* cell wall by ELISA.

Next, transmission electron microscopy was performed to examine if there were any differences in the cell wall structure and composition between wild type, *ΔpknF*, *ΔRv1747* and the respective complementing strains. In all cases there were no discernable differences in cell wall structure between the strains. **This result showed that Rv1747 does not appear to export any component of the cell wall that is involved in formation of an observable structure otherwise a difference would have been noted between the strains in the microscopy**.

Protein composition of the supernatants of wild type, *ΔpknF*, *ΔRv1747*, and the respective complementing strains were examined throughout growth by SDS-PAGE. **No differences in protein composition were observed between the strains**. One could hypothesise that there might have been a protein absent in the *ΔRv1747* strain if the ABC transporter did in fact export a protein. However, this result suggested that Rv1747 does not export a protein. Nevertheless, perhaps this technique is not sensitive enough to spot any protein differences between the strains. One alternative technique that could be attempted is two dimensional (2D) gel electrophoresis. This technique separates all proteins in a sample based on two properties, for example protein mass and isoelectric point. Protein molecules are more effectively separated in this system compared with the SDS-PAGE performed in this study as proteins are much less likely to share two different properties.

Nuclear magnetic resonance spectroscopy was performed to investigate whether there were any differences in the small molecules present in the culture supernatants of the wild type, *ΔRv1747* and *Rv1747* complementing strains. The hypothesis was that Rv1747 could export a small molecule or metabolite which is essential for the virulence of *M. tuberculosis*. Any product or intermediate from metabolism is defined as a metabolite. For example this could include amino acids, chemical messengers such as cyclic AMP and other signalling molecules. Results showed that yet again there were no differences in the small molecules present between the strains meaning that the **Rv1747 transporter probably does not export anything detectable by NMR**. However, if the substrate of Rv747 is incorporated into part of the *M. tuberculosis* cell then NMR would not have detected it as this technique only examined molecules present in the culture supernatants.

Finally, Phenotype MicroArrays^{TM} were used to assess if either of the mutants had a phenotype when subjected to various substrates. Prior to inoculation the *M. tuberculosis* cultures were washed and then starved in PBS for 18 hours to ensure that the cells had metabolised any nutrients, carbon sources and other growth factors remaining from the Dubos medium. This was important as otherwise false positive results may have been observed. It was hypothesised that any phenotypic differences observed between the

strains may help in identifying a function for PknF and Rv1747. However, results showed that **there were no reproducible differences in substrate usage found between the wild type and mutant strains**. Although disappointing in this study, Phenotype MicroArray^{TM} technology has been successful in previous investigations to help identify substrates for proteins and to help elucidate their function. For example, a study in 2006 showed that Phenotype MicroArray^{TM} technology was an effective screening method to assess cellular phenotypes in *Staphylococcus aureus* (von Eiff *et al.*, 2006). von Eiff and colleagues were able to characterise the phenotype of two mutants using this technique, namely *menD* and *hemB*, both of which form small colonies. They discovered that both mutants had defects in carbon utilisation and carbon metabolism and these were confirmed in further growth studies. In addition they were able to record digital images of the microarrays while they were in the incubator which stored colour change data thus generating real-time kinetics results. This latter technique is not currently available in our laboratory but highlights the potential of the method to screening *M. tuberculosis* mutants with no known phenotype.

In conclusion, the results presented in this chapter have highlighted the inherent difficulties of trying to identify the substrate for Rv1747. Although many hypotheses were tested, the experiments performed did not yield conclusive evidence as to the nature of the substrate. Furthermore, although lipid profile analysis highlighted a difference in lipid composition between wild type and mutant strains this difference could not be complemented. Consequently, **the identity of the Rv1747 substrate has remained elusive**. In the future if a potential substrate is identified this can be tested using inverted membrane vesicles in a transport assay. Briefly, inverted membrane vesicles are enriched with Rv1747, then in the presence of ATP the substrate in transported inside the inverted vesicles and can then be quantified.

In further chapters of this study the biochemical aspects of protein-protein interactions will be explored and the mechanism of regulation will be investigated in order to further understand how PknF regulates the function of Rv1747.

CHAPTER 5. HOW IS PknF CONTROLLING Rv1747 FUNCTION? PART 1

ANALYSIS OF THE PHOSPHORYLATION DEPENDENT INTERACTIONS BETWEEN RECOMBINANT PknF AND Rv1747 *IN VITRO*

5.1 Introduction

Serine-threonine protein kinases (STPKs) are key proteins in signalling systems which act by regulating the function of their protein substrates. Eleven STPKs have been identified in the *M. tuberculosis* genome, from PknA to PknL (there is no PknC), nine of which are predicted to be transmembrane bound kinases (Av-Gay & Everett, 2000; Cole *et al.*, 1998). This study focussed on the transmembrane bound kinase PknF with the aim of exploring how this kinase was controlling the function of Rv1747, an FHA domain containing ABC transporter required for the full virulence of *M. tuberculosis* (Curry *et al.*, 2005). The hypothesis was that PknF regulates the function of Rv1747 in a phosphorylation dependent manner. As it was unknown how PknF was controlling Rv1747 function, this study initially characterised the phosphorylation dependent interactions between the two recombinant proteins.

Numerous studies have been conducted which have characterised the interactions between other *M. tuberculosis* STPKs and their putative FHA domain containing substrates (Molle *et al.*, 2003; Nott *et al.*, 2009; O'Hare *et al.*, 2008; Villarino *et al.*, 2005). However, in the case of PknF and Rv1747 only two studies have been published which presented a preliminary analysis of the interaction between these two proteins (Curry *et al.*, 2005; Molle *et al.*, 2004). In 2005 Curry *et al.* used yeast two-hybrid analysis to demonstrate that the Rv1747 protein interacted with PknF (Curry *et al.*, 2005). Furthermore, they went on to demonstrate that the interaction appeared to be phosphorylation dependent as the interaction was abrogated in a kinase null mutant, when the presumed activation loop of PknF was mutated and when the FHA-1 domain of Rv1747 was mutated (Curry *et al.*, 2005). In 2004 Molle *et al*. used *in vitro* phosphorylation assays to show that the PknF kinase domain phosphorylated the Rv1747

nucleotide binding domain *in vitro*. However, in their purification of Rv1747 they kept the large GST-fusion tag as part of the recombinant protein. Therefore it could be argued firstly that GST-Rv1747 was not folded correctly, due to the large tag, and therefore serine or threonine residues not normally accessible to PknF were now rendered accessible to phosphorylation and vice versa, and secondly, the kinase may have phosphorylated a residue within the GST protein itself. Therefore in this current investigation all recombinant proteins were produced without any tags unless necessary for a specific technique.

Both these studies highlighted that there was probably a phosphorylation dependent interaction between PknF and Rv1747; however, the identity of the phosphorylation sites and the importance of these sites on the ability of PknF to phosphorylate Rv1747 and regulate its activity had not been explored. This therefore warranted further investigation to help gain a more in depth understanding of how these two proteins interacted and how PknF controlled Rv1747 function in the context of *M. tuberculosis* infection and virulence. The consequences of phosphorylation of Rv1747 by PknF will be investigated in this chapter and the following chapter of this study.

The main objective of this study was to begin to understand the mechanism by which PknF controls the function of Rv1747. It was hypothesised that PknF controlled the function of Rv1747 by phosphorylation on specific threonine residues. The first aim of this study was to express and purify PknF and Rv1747 from *E. coli*. The second aim was to ascertain whether recombinant PknF was an active kinase and to identify whether this kinase could phosphorylate Rv1747 *in vitro*; the next objective was to identify the specific sites on Rv1747 which were phosphorylated by PknF. Finally, the significance of these sites were characterised in *in vitro* phosphorylation assays.

5.2 Expression and purification of PknF and Rv1747 from *E. coli*

As a necessary preliminary to exploring the phosphorylation dependent interactions between PknF and Rv1747 both the recombinant proteins were expressed and purified from *E. coli*.

The kinase domain of PknF and the nucleotide binding domain of Rv1747 were cloned into the N-terminal GST-fusion vector pGEX-6P-1 and then expressed and purified from BL21 star *E.coli* cells. Figure 5.1 highlights the domain architecture of PknF and Rv1747 according to the SMART database (Schultz *et al.*, 1998). Figure 5.1a shows the domain architecture of PknF which comprises of a kinase domain (shown in red) and a transmembrane domain (shown in blue). It was important not to include any of the transmembrane domains in the protein expression constructs as it is inherently difficult to produce soluble proteins which contain these domains due to their hydrophobicity. Furthermore, the aim of the investigation was to study the function of the kinase domain so the transmembrane domain was not required. According to the SMART database the PknF kinase domain ends at amino acid 274 and the transmembrane domain begins at residue 307. Therefore amino acid residues 1 to 292 were chosen to be cloned into pGex-6P-1 to ensure the complete kinase domain was present but to make sure the transmembrane domain was not included (Figure 5.1a). In figure 5.1a this region is highlighted with a black box. The protein domains of Rv1747 are depicted in Figure 5.1b. According to the SMART database the nucleotide binding domain ends at amino acid 528 and the transmembrane domain begins at position 613. Therefore amino acids 1 to 559 were cloned into pGex-6P-1 for the Rv1747 expression work to ensure that the complete ATPase domain was included in the construct and to make sure the first transmembrane domain was excluded. In figure 5.1b this region is highlighted with a black box.

PknF$^{1\text{-}292}$ and Rv1747$^{1\text{-}559}$ were PCR amplified from *M. tuberculosis* genomic DNA, ligated into pGex-6-P1 and confirmed by sequencing as described in sections 2.2.2, 2.2.9 and 2.2.13 respectively. These DNA fragments were then transformed into either *E. coli* BL21 star (DE3) cells in the case of PknF$^{1\text{-}292}$ or *E. coli* BL21 star (DE3) pRep4-*groESL* cells in the case of Rv1747$^{1\text{-}559}$. Initial solubility and purification trials found that it was very difficult to identify an expression condition which yielded a high abundance of soluble Rv1747$^{1\text{-}559}$ protein in BL21 star cells which could then be purified from its contaminants (data not shown). An alternative expression cell type for Rv1747$^{1\text{-}559}$ purification was then sought. The pRep4-*groESL* plasmid encodes GroEL and GroES, two bacterial molecular chaperones (Amrein *et al.*, 1995). The yield of soluble

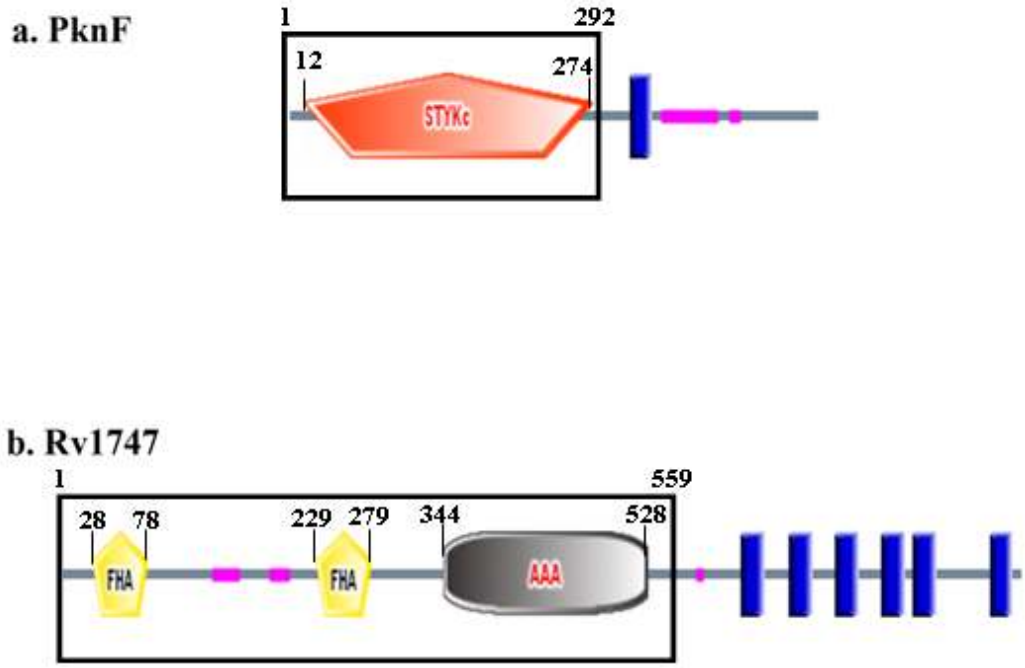


Figure 5.1 Protein domains cloned for expression and purification from *E. coli*. Figure 5.1a shows the domain architecture of PknF and figure 5.1b shows the domain architecture of Rv1747. Images were produced using the SMART database (Letunic *et al.*, 2009). Figure 5.1a highlights the kinase domain of PknF (in red) plus the transmembrane (blue) and regions of low complexity (pink). Figure 5.1 b shows both the FHA domains of Rv1747 (yellow), the ATPase domain (grey), the transmembrane domains (blue) and the regions of low complexity (pink). Numbers represent amino acid positions. Cloned residues are highlighted with a black box.

Rv1747$^{1-559}$ protein in the BL21 star (DE3) pRep4-*groESL* expression strain was much higher and it was possible to purify the soluble protein from its contaminants using size exclusion chromatography. This *E. coli* expression strain was therefore used for all Rv1747$^{1-559}$ purifications.

PknF$^{1-292}$ and Rv1747$^{1-559}$ were expressed and purified as described in section 2.5.2. The N-terminal GST-tag was removed from both the kinase and transporter during the purification using PreScission protease. It was felt that it was important to remove the tag from both the proteins for two reasons; firstly, the tag is comparatively large (~26 kDa) and so it was thought that this may interfere with the correct folding of the two recombinant proteins; if the proteins did not fold correctly then it would be likely that they would be rendered non-functional. Secondly, as Rv1747 was to be used in phosphorylation assays to determine if and where PknF phosphorylated the ABC transporter, it was decided that the only domains present in the Rv1747 purified protein would be those from Rv1747 itself. Cleavage of the tag from the proteins left five amino acids at the N-terminus of each protein: GPLGS.

Figure 5.2 shows the results of the purifications of both the kinase and the ABC transporter. All protein purifications required a size exclusion chromatography step to remove contaminating proteins that eluted after PreScission protease cleavage. Figure 5.2a is a UV trace generated from size exclusion chromatography of a PknF$^{1-292}$ purification. Size exclusion chromatography was performed as outlined in section 2.5.2.2.1. Figure 5.2a highlights the UV absorbance peak corresponding to PknF$^{1-292}$. All other UV absorbance peaks are from contaminating proteins which separated into other fractions. When both the proteins had been separated from their contaminants by size exclusion chromatography the fractions containing the desired protein were pooled and concentrated. 10 µl of this sample was then analysed by SDS-PAGE; figure 5.2b shows the result of a PknF$^{1-292}$ purification and figure 5.2c shows the result of a Rv1747$^{1-559}$ purification. Figure 5.2b shows that PknF$^{1-292}$ was purified to homogeneity. Interestingly this protein also runs as a doublet. The identity of each purified protein was confirmed by mass spectrometry and was performed as described in 2.5.2.3. Figure 5.2c shows that Rv1747$^{1-559}$ was purified to near homogeneity; the contaminating protein at

approximately 70 kDa was identified by mass spectrometry as *E. coli* heat shock protein 70. This chaperone protein was always present in $Rv1747^{1-559}$ purifications.

5.3 Recombinant $PknF^{1-292}$ is an active kinase capable of autophosphorylation

In order to determine whether recombinant $PknF^{1-292}$ was a functional kinase, an *in vitro* phosphorylation assay was performed where $PknF^{1-292}$ was labelled with [γ-^{32}P] ATP for 30 minutes before being separated by gel electrophoresis (figure 5.3a) and checked for incorporation of radioactivity by autoradiography (figure 5.3b). Figure 5.3b lane 2 clearly shows that $PknF^{1-292}$ incorporated radio-labelled ATP demonstrating that it was indeed an active kinase capable of autophosphorylation. Furthermore, electrospray mass spectroscopy confirmed the presence of additional phosphate groups on the recombinant kinase after purification from *E. coli* indicating that *in vitro* $PknF^{1-292}$ is capable of autophosphorylating itself numerous times (figure 5.3c).

5.4 Autophosphorylation activity of $PknF^{1-292}$ is dependent on one specific lysine residue

Archetypal serine-threonine protein kinases contain a lysine residue within their active site that is critical for autophosphorylation due to its role in ATP binding (see figure 1.4) (Kornev *et al.*, 2006; Young *et al.*, 2003). To further explore whether this was the case for $PknF^{1-292}$ site directed mutagenesis was employed to make a K41A mutation within the expression construct. The result of the $PknF^{1-292}$ K41A purification is shown in figure 5.4a. $PknF^{1-292}$ K41A was then assessed for its autophosphorylation capabilities using an *in vitro* kinase assay. Figure 5.4c lane 2 shows that the K41A mutation abolished the autophosphorylation activity of $PknF^{1-292}$ which was observed in the wild type control in lane thus confirming the essentiality of the lysine residue for kinase function. More importantly, this result demonstrated that the autophosphorylation activity of $PknF^{1-292}$ observed in the *in vitro* phosphorylation experiment was the result of autocatalytic activity and not due to phosphorylation by an *E. coli* protein during the purification process.

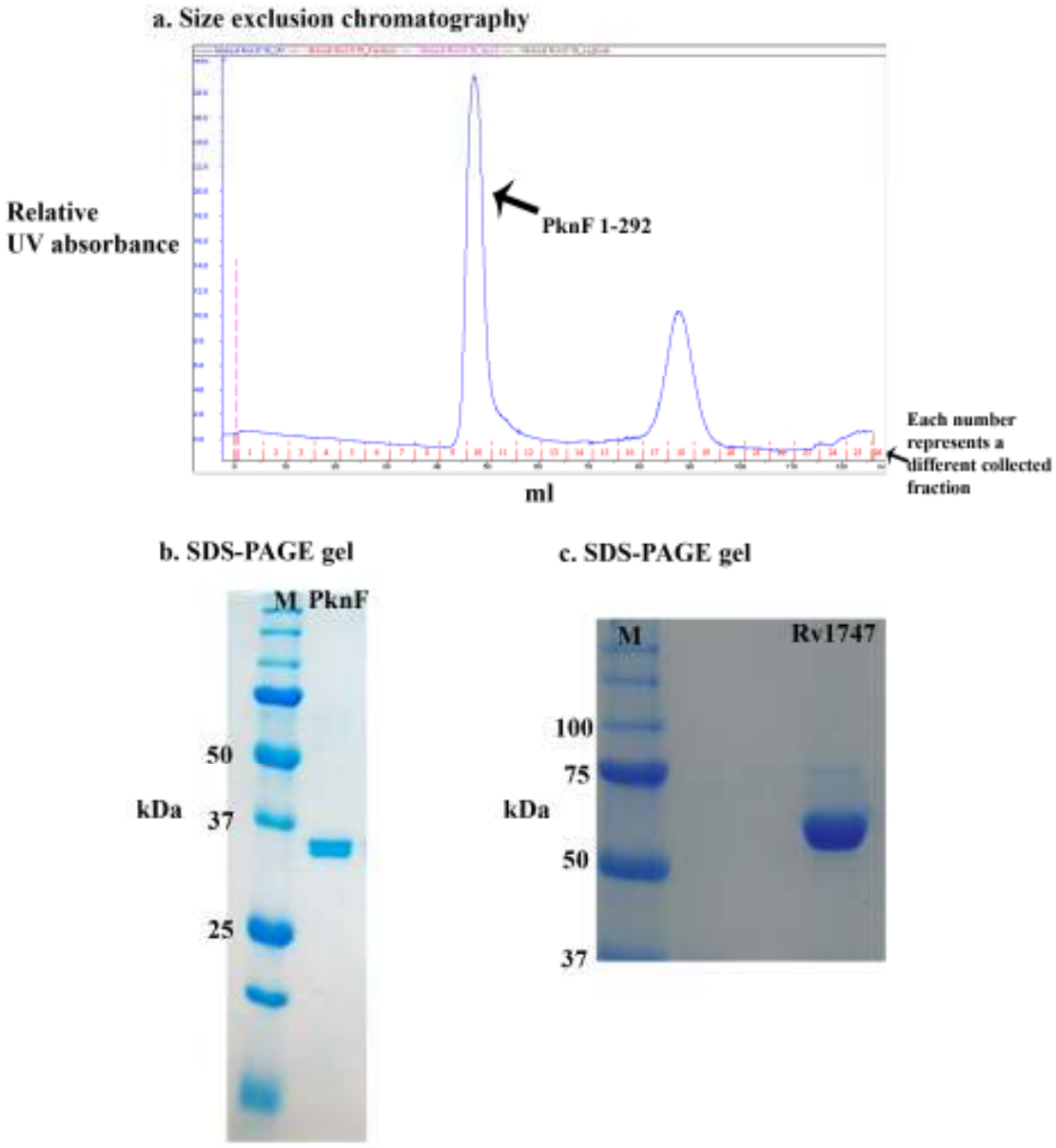
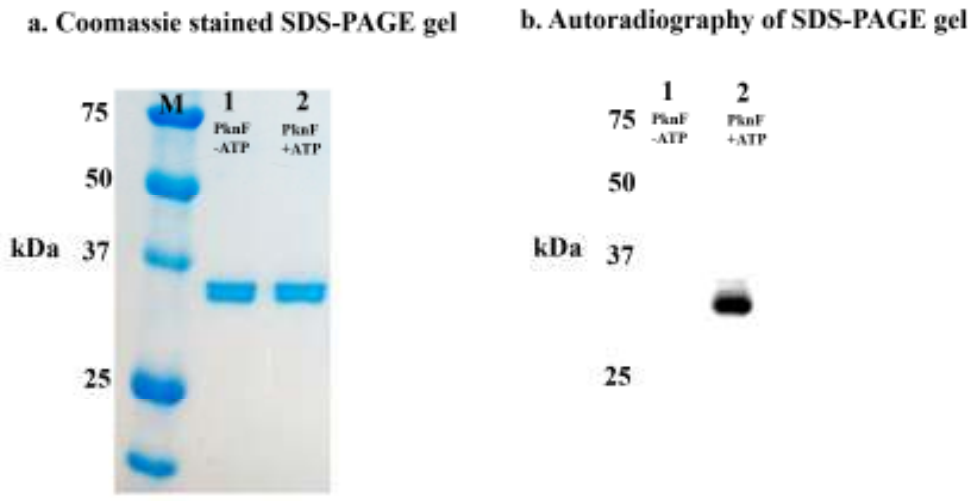


Figure 5.2 Expression and purification of recombinant PknF$^{1\text{-}292}$ and Rv1747$^{1\text{-}559}$ from *E.coli*. PknF$^{1\text{-}292}$ and Rv1747$^{1\text{-}559}$ were expressed and purified as described in section 2.5.2. All protein purifications required a size exclusion chromatography step to remove contaminating proteins. Size exclusion chromatography was performed with a gel filtration column (HiLoad 16/60 Superdex 200 prep grade column) and an AKTA Prime system as described in section 2.5.2.2.1. Figure 5.2a shows a UV trace from a PknF$^{1\text{-}292}$ purification. All other UV absorbance peaks are from contaminating proteins which separated into other fractions. After purification 10 µl of each pooled and concentrated protein sample was analysed by SDS-PAGE as described in section 2.5.1. Figure 5.2b shows the result of a PknF$^{1\text{-}292}$ purification and figure 5.2c shows the results of an Rv1747$^{1\text{-}559}$ purification. PknF$^{1\text{-}292}$ was purified to homogeneity. Rv1747$^{1\text{-}559}$ was purified to near homogeneity; the contaminating protein at ~ 70kDa was identified as *E. coli* heat shock protein 70. This contaminating protein was always present in all Rv1747$^{1\text{-}559}$ purifications. PknF$^{1\text{-}292}$ = 32 kDa, Rv1747$^{1\text{-}559}$ = 58 kDa.



c. Mass spectrometry of PknF 1-292

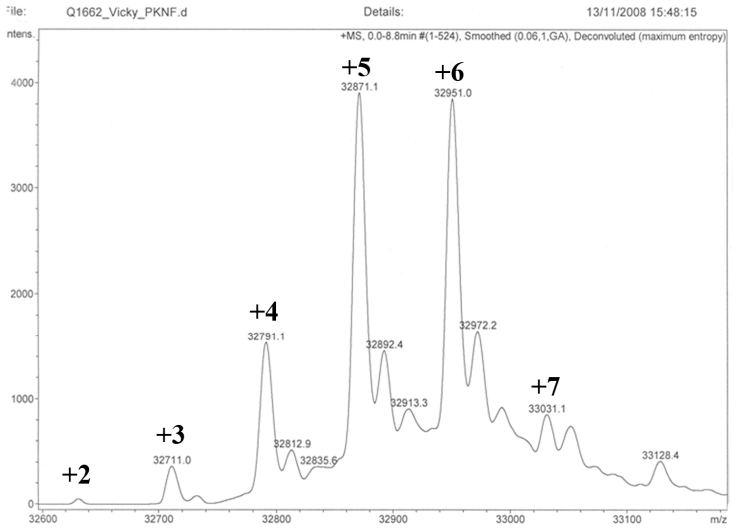


Figure 5.3. Recombinant PknF$^{1-292}$ is an active kinase capable of autophosphorylation. *In vitro* phosphorylation assays were performed with expressed and purified PknF$^{1-292}$ as described in section 2.5.3.1. PknF$^{1-292}$ was incubated for 30 minutes at 37 ^{o}C, in the case of lane 2, in the presence of 200 µCi/ml [γ-^{32}P] ATP, and subjected to gel electrophoresis (Figure 5.2a). Radioactive bands were revealed by autoradiography (Figure 5.2b). Approximately 1 µg of protein was loaded into each lane. Figure 5.3b lane 2 shows a radioactive signal corresponding to the size of PknF$^{1-292}$ (32.4 kDa) demonstrating that the kinase is capable of autophosphorylation. Figure 5.3c shows an electrospray mass spectrometry analysis of PknF$^{1-292}$ showing the autophosphorylation activity of the PknF$^{1-292}$ kinase domain. Each peak labeled '+' followed by a number represents the number of additional phosphate groups on PknF$^{1-292}$; i.e. '+2' means PknF $^{1-292}$ plus 2 phosphate groups. The net gain of a phosphate group results in an increase in mass of 80 Da.

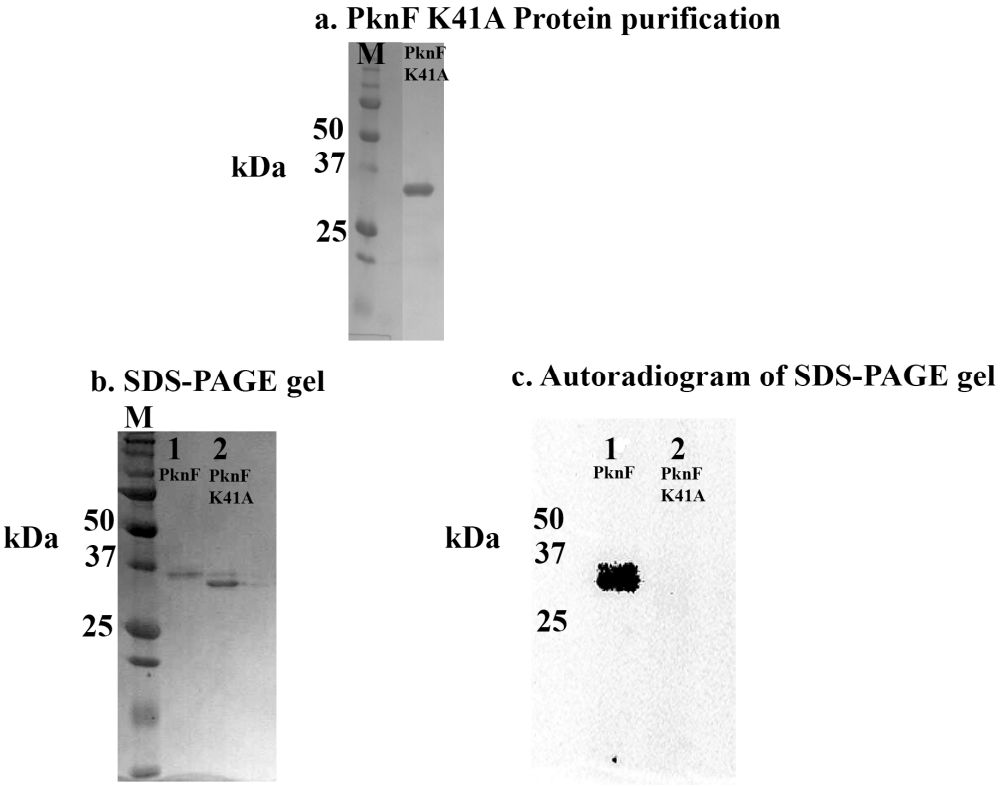


Figure 5.4 A mutation of the critical lysine residue abolishes the autophosphorylation activity of PknF$^{1-292}$. Site directed mutagenesis was used generate a K41A mutation within the PknF$^{1-292}$ construct as described in section 2.2.3. Figure 5.4a shows that PknF$^{1-292}$ K41A was expressed and purified to homogeneity. *In vitro* phosphorylation assays were performed with wild type PknF$^{1-292}$ and PknF$^{1-292}$ K41A. The proteins were incubated with 200 µCi/ml [γ-^{32}P] ATP for 30 minutes, subjected to gel electrophoresis (Figure 5.3b) and radioactive bands were revealed by autoradiography (Figure 5.4c). Approximately 0.5 µg of protein was loaded into each lane. Figure 5.4c lane 1 shows that wild type PknF$^{1-292}$ is capable of autophosphorylation. Figure 5.4c lane 2 clearly demonstrates that the autophosphorylation activity of the kinase was abolished in the PknF$^{1-292}$ K41A mutant.

5.5 PknF$^{1-292}$ phosphorylates Rv1747$^{1-559}$ *in vitro*

Prior to any experiments being carried out to determine the sites of Rv1747 phosphorylation it first had to be determined whether PknF could phosphorylate Rv1747 *in vitro*. Both the recombinant proteins were combined in an *in vitro* phosphorylation assay. Figure 5.5b lane 2 shows that PknF$^{1-292}$ does indeed phosphorylate Rv1747$^{1-559}$ *in vitro* as demonstrated by the presence of a radio-labelled band corresponding to the size of Rv1747$^{1-559}$ (58 kDa). Furthermore, as expected, Rv1747 does not have any autophosphorylation activity (figure 5.5d lane 1).

5.6 PknF phosphorylates Rv1747 *in vitro* on specific threonine residues

The next aim of this study was to identify the sites on Rv1747 which were phosphorylated by PknF. Phosphorylation site mapping was achieved using LC-MS/MS as described in section 2.5.3.2. Purified PknF$^{1-292}$ and Rv1747$^{1-559}$ were supplied to V. Molle (Lyon, France) for site identification. Two phosphorylated threonine residues were identified on Rv1747$^{1-559}$ and these are shown in figure 5.6. The two phosphorylated threonine residues identified were T150 and T208. T150 is situated between the FHA-1 domain and the FHA-2 domain and T208 is just upstream of the FHA-2 domain core region (figure 5.6).

5.7 Construction of Rv1747 phosphorylation-null expression strains in *E. coli*

It was hypothesised that PknF controlled the function of Rv1747 through phosphorylation on specific threonine residues. To further investigate the importance of the identified pT150 and pT208 sites on the ability of PknF to phosphorylate Rv1747 *in vitro* a panel of threonine to alanine mutations were created in the Rv1747$^{1-559}$ protein expression construct. Three mutations were generated; T150A, T208A and a double T150A/T208A mutant. Replacement of a threonine residue with an alanine residue renders the site incapable of acting as a phosphoacceptor (phosphorylation-null). Site directed mutagenesis was performed as described in section 2.2.3 and a list of primers can be found in appendix II. Two rounds of site directed mutagenesis were required to

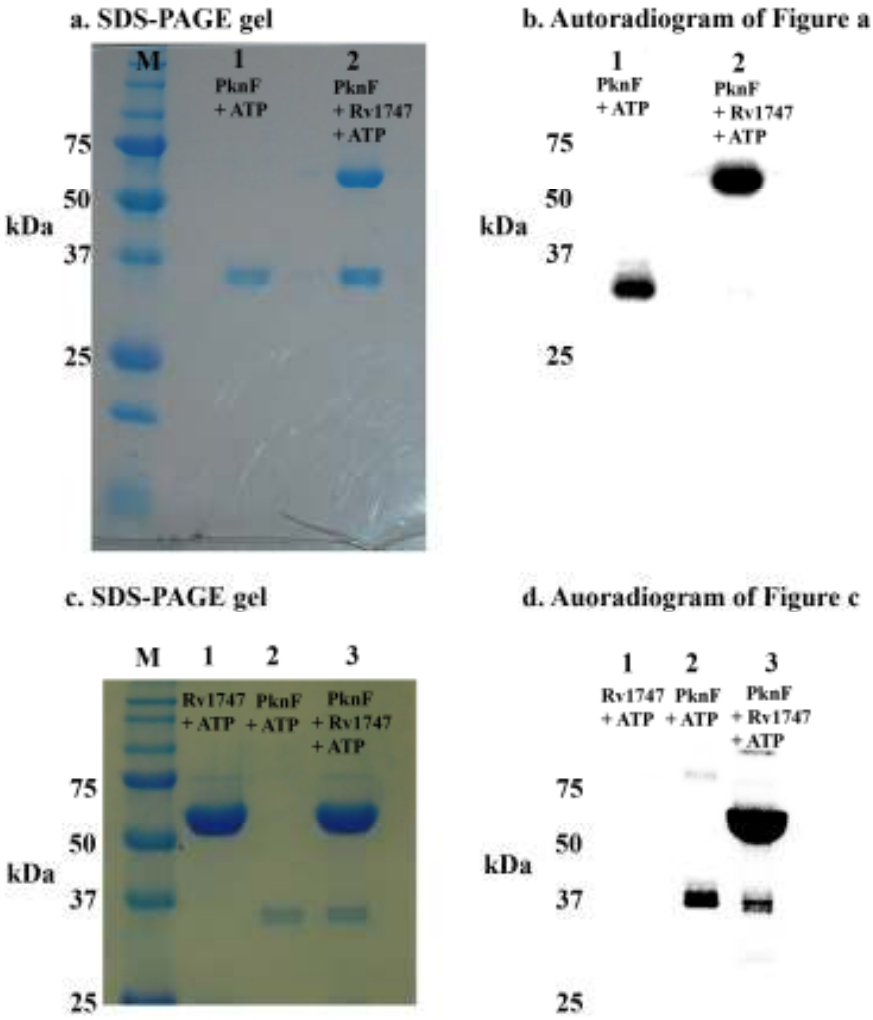


Figure 5.5 *In vitro* phosphorylation of Rv1747 by PknF. *In vitro* phosphorylation assays were performed with $PknF^{1-292}$ and $Rv1747^{1-559}$ as outlined in section 2.5.3.1. The proteins were incubated with 200 µCi/ml [γ-^{32}P] ATP for 30 minutes, subjected to gel electrophoresis (Figure 5.5a and c) and radioactive bands were revealed by autoradiography (Figure 5.5b and d). Approximately 1 µg of protein was loaded into each lane. Figure 5.5b shows that $PknF^{1-292}$ is an active kinase capable of autophosphorylation (lane 1) and that it is able to phosphorylate $Rv1747^{1-559}$ *in vitro* (lane 2). Figure 5.5d shows that Rv1747 does not have any autophosphorylation activity (lane 1) and the relevant controls are shown in lanes 2 and 3. $PknF^{1-292}$ = 32 kDa, $Rv1747^{1-559}$ = 58 kDa.

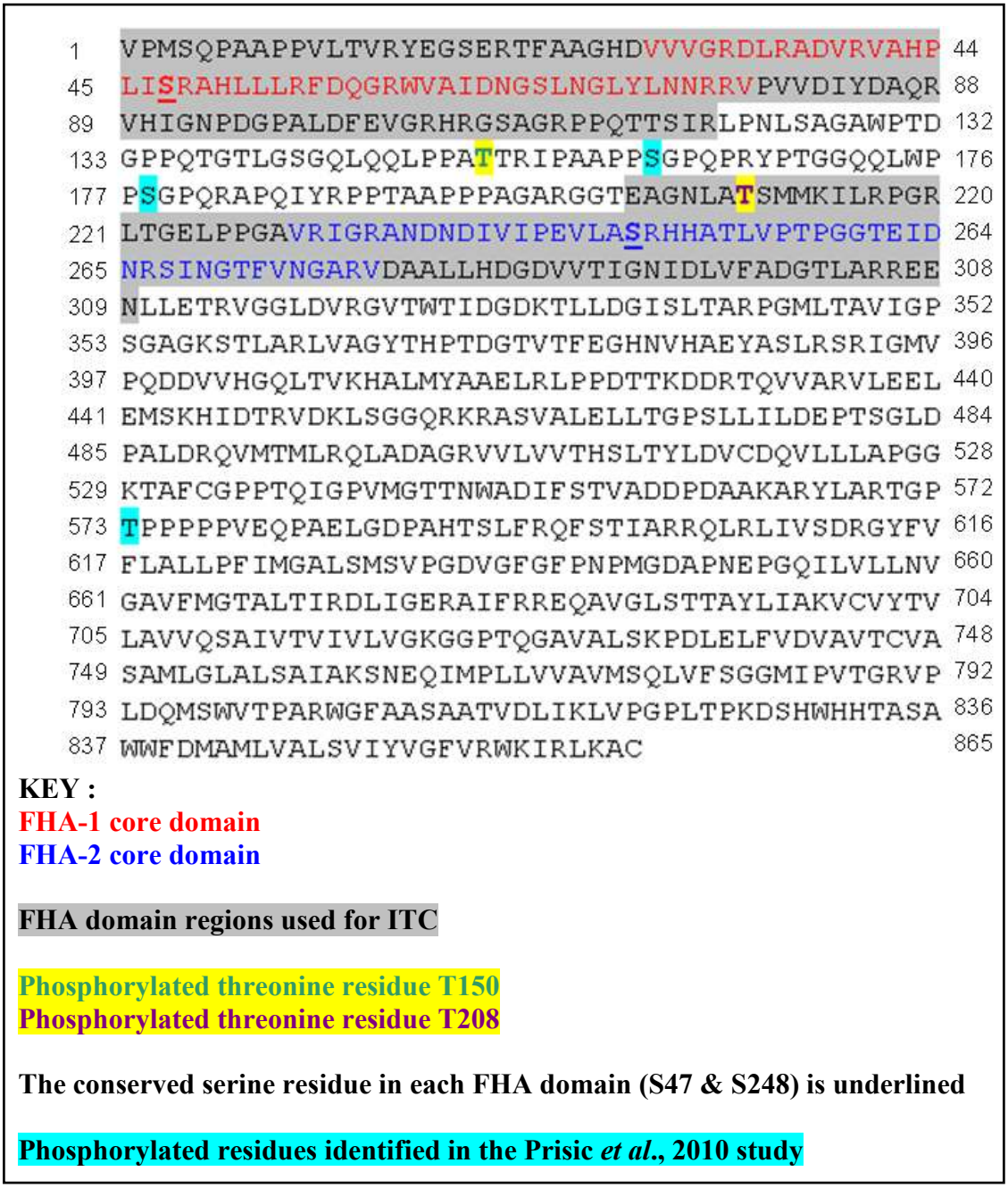


Figure 5.6 Identification of phosphorylated threonine residues on Rv1747$^{1-559}$. Recombinant PknF$^{1-292}$ and Rv1747$^{1-559}$ were expressed and purified as outlined in section 2.5.2. LC/MS/MS was performed in collaboration with V. Molle as described in section 2.5.3.2 to identify phosphorylated threonine residues. Two residues were discovered which were phosphorylated *in vitro* by PknF$^{1-292}$. The two residues are highlighted in figure 5.6 which shows Rv1747 amino acid residues 1 to 559. The two phosphorylated threonine residues identified were T150 and T208 which are highlighted in green and purple respectively. The two FHA domains are also highlighted in figure 5.6; the FHA-1 core domain is highlighted red and the FHA-2 core domain is highlighted in blue. The amino acids highlighted in grey show the residues cloned to make the FHA-1 (pVS_04) and FHA-2 (pVS_05) protein expression constructs.

generate the double threonine to alanine mutant. It was expected that a reduction in phosphorylation would be observed in the double threonine to alanine mutant if indeed the two identified sites were important for Rv1747 phosphorylation by PknF.

The effects of the threonine to alanine mutations on the ability of $PknF^{1-292}$ to phosphorylate $Rv1747^{1-559}$ *in vitro* were then assessed.

5.8 Expression and purification of $Rv1747^{1-559}$ threonine to alanine mutants in *E. coli*

Figure 5.7 shows the results of the expression and purification of the threonine to alanine mutant proteins from *E. coli*. 10 µl of each concentrated protein sample was analysed by SDS-PAGE. It can be seen that all proteins were purified to near homogeneity with the exception of the contaminating *E. coli* hsp70 protein which was present in all $Rv1747^{1-559}$ purifications.

5.8.1 Assessment of the phosphorylation state of the $Rv1747^{1-559}$ mutants

Next, *in vitro* phosphorylation assays were performed to examine the effects of the threonine to alanine mutations on the ability of PknF to phosphorylate Rv1747. Results of the assays are shown in figure 5.8. After incubation with [γ-^{32}P] ATP proteins were subjected to gel electrophoresis (figure 5.8a) and radioactive bands were revealed by autoradiography (figure 5.8b). Phosphorylation of $Rv1747^{1-559}$ by $PknF^{1-292}$ is reduced in all the threonine to alanine mutants (figure 5.8b) and was most significantly reduced in the double threonine to alanine mutant (figure 5.8b, lane 5). This result demonstrated that phosphorylation of Rv1747 by PknF is highly impaired in a double threonine to alanine mutant in sites T150 and T208 of Rv1747 *in vitro*. If the two phospho-threonine sites identified were the *only* sites on Rv1747 phosphorylated by PknF then it was expected that in the double mutant no radioactive band corresponding to the phosphorylation of Rv1747 would be observed. Interestingly, although very little radio labelled ATP was incorporated into the Rv1747 protein in the double threonine to alanine mutant, there also was no radioactive band present corresponding to the size of the kinase. This was a very surprising result as one would expect the kinase to remain

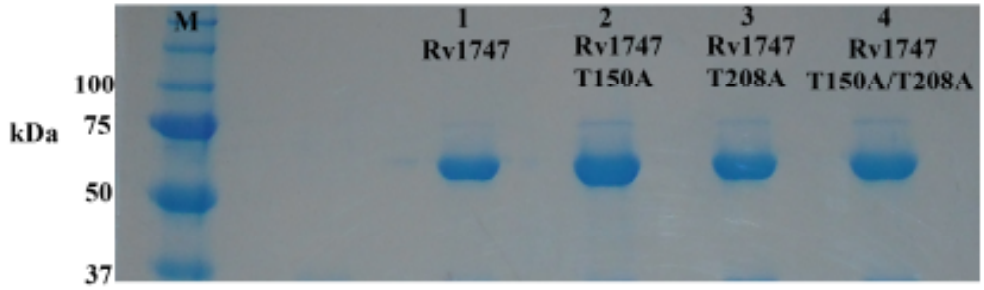


Figure 5.7 Expression and purification of wild type Rv1747$^{1-559}$, Rv1747$^{1-559}$ T150A, Rv1747$^{1-559}$ T208A and Rv1747$^{1-559}$ T150A/T208A. Proteins were expressed and purified as described in section 2.5.2. After purification 10 µl of each pooled and concentrated protein sample was analysed by SDS-PAGE as described in section 2.5.1. Figure 5.7 demonstrates that all proteins were purified to near homogeneity; the contaminating protein at ~ 70kDa was identified as *E. coli* heat shock protein 70. The contaminating protein was always present in all Rv1747$^{1-559}$ purifications. Rv1747$^{1-559}$ = 58 kDa.

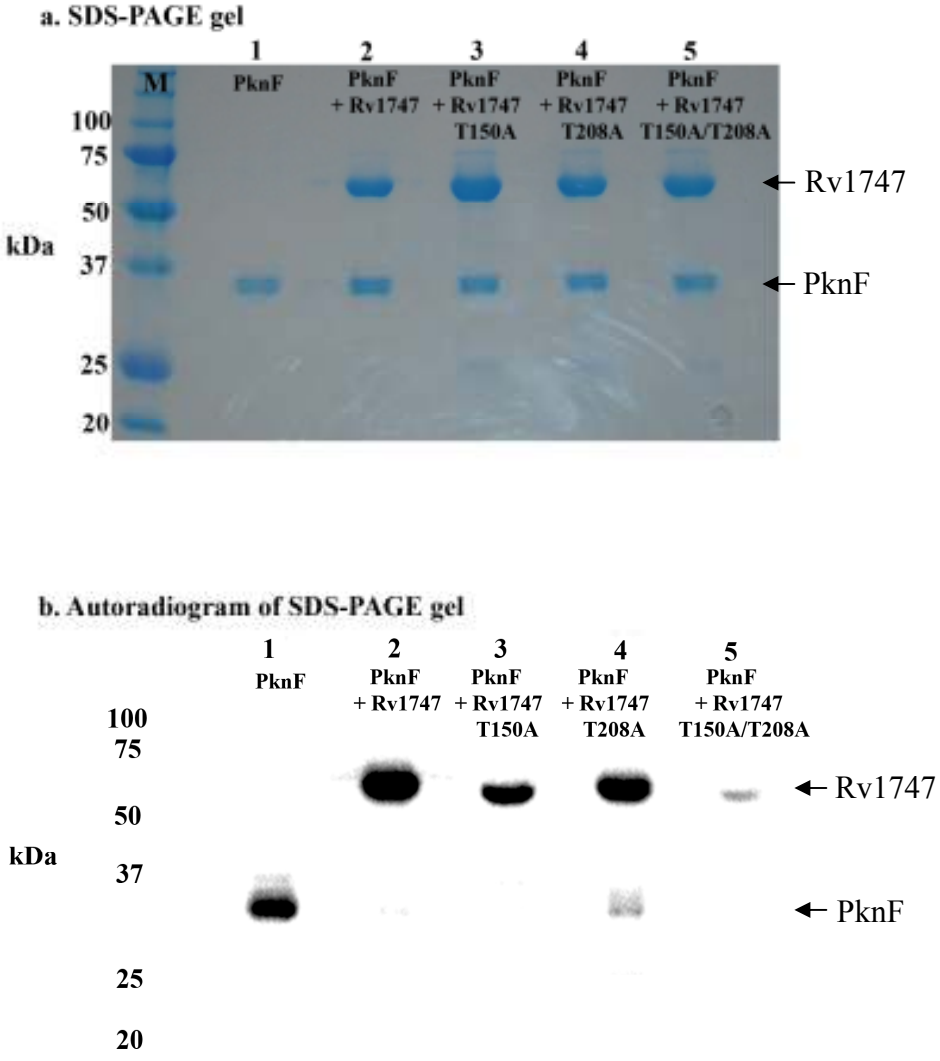


Figure 5.8 *In vitro* phosphorylation of Rv1747$^{1-559}$ by PknF$^{1-292}$ is reduced in the threonine to alanine mutants. *In vitro* phosphorylation assays were performed with PknF$^{1-292}$ and wild type Rv1747$^{1-559}$, Rv1747$^{1-559}$ T150A, Rv1747$^{1-559}$ T208A or Rv1747$^{1-559}$ T150A/T208A as outlined in section 2.5.3.1. The various proteins were incubated together with 200 µCi/ml [γ-^{32}P] ATP for 30 minutes, subjected to gel electrophoresis (Figure 5.8a) and radioactive bands were revealed by autoradiography (Figure 5.8b). Approximately 1 µg of protein was loaded into each lane.

autophosphorylated in this situation when it could no longer phosphorylate Rv1747 sufficiently.

5.9 Assessment of the phosphorylation of Rv1747$^{1-559}$ by other *M. tuberculosis* serine-threonine protein kinases

It has long been hypothesised that more then one *M. tuberculosis* serine-threonine protein kinase could phosphorylate the same substrate *in vivo* leading to a level of redundancy between the kinases and suggesting that substrates may be regulated by multiple signals (Grundner *et al.*, 2005). This hypothesis is further supported by the fact that only two out of the 11 STPKs have to date been shown to be essential for *M. tuberculosis* viability (Sassetti *et al.*, 2003). To test this hypothesis with Rv1747 in our *in vitro* model, phosphorylation assays were performed as previously with Rv1747, PknF, PknG and PknB. Figure 5.9 shows the results from the assays. The radioactive bands present in lanes 1 and 3 in figure 5.9b demonstrate that both the kinases were active and capable of autophosphorylation. Lane 2 shows the Rv1747$^{1-559}$ control demonstrating that Rv1747 was phosphorylated by PknF$^{1-292}$. Figure 5.9b lane 4 shows that PknG does not phosphorylate Rv1747 *in vitro* demonstrated by the absence of a radio-labelled band corresponding to the size of Rv1747. Figure 5.9c lane 1 clearly demonstrates that PknB was a functionally active kinase but Rv1747 was not a good substrate for this kinase. This result clearly illustrated that PknG could not phosphorylate Rv1747$^{1-559}$ and that PknB could only marginally phosphorylate Rv1747$^{1-559}$ *in vitro*. Again, it was interesting that the autophosphorylation activity of PknG and PknB was dramatically reduced when the Rv1747 protein was included in the phosphorylation assay compared with the levels of ATP incorporated into the kinases when they were incubated in the absence of Rv1747.

As a control experiment to check that PknG$^{1-759}$ and PknB$^{1-279}$ were indeed functional kinases it was investigated whether they could phosphorylate Rv1827. Rv1827 is another FHA domain containing protein in *M. tuberculosis* which is phosphorylated by PknB and PknG on two specific threonine residues (Nott *et al.*, 2009; O'Hare *et al.*, 2008; Villarino *et al.*, 2005). The results from the *in vitro* phosphorylation assays are

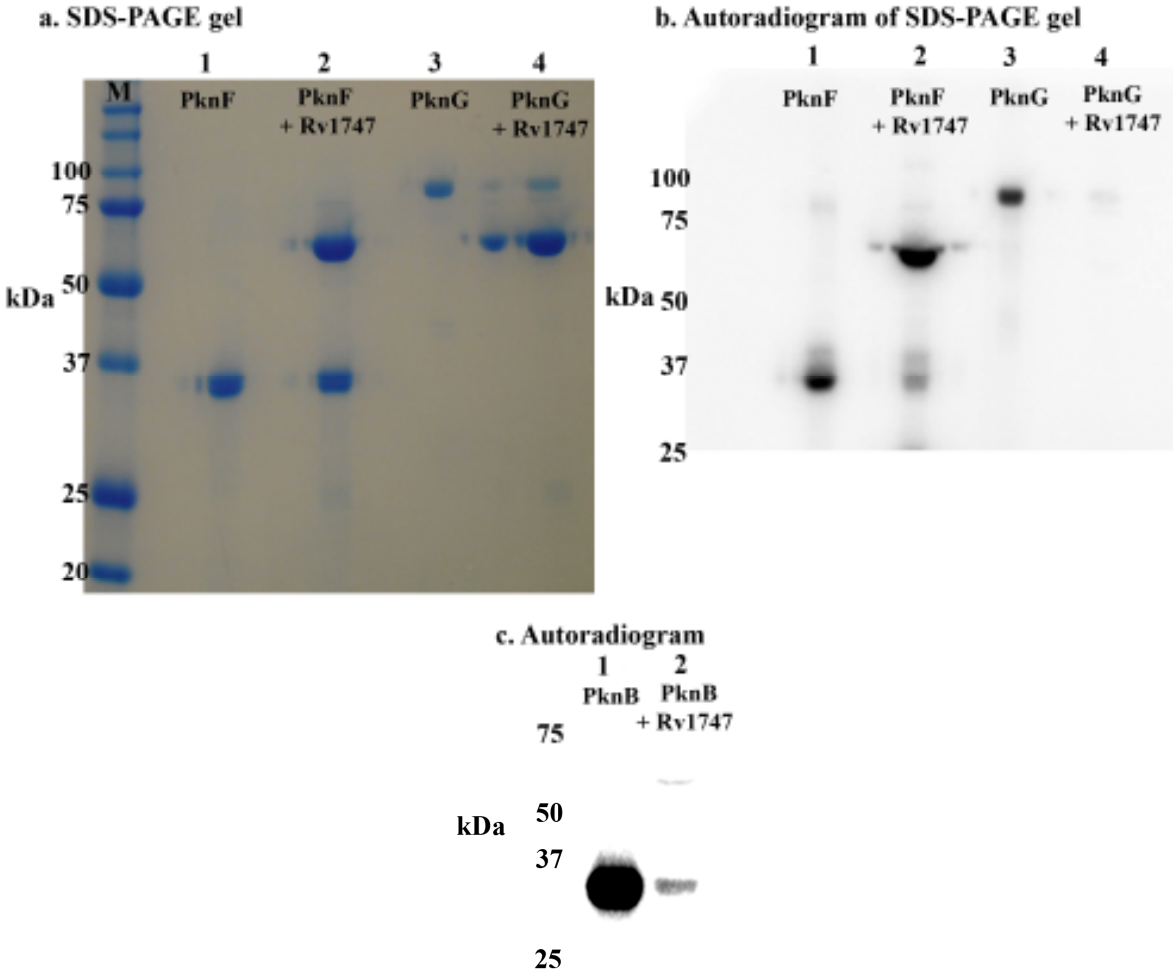


Figure 5.9 Assessment of the phosphorylation of Rv1747$^{1-559}$ by other *M. tuberculosis* serine-threonine protein kinases. *In vitro* phosphorylation assays were performed with wild type Rv1747$^{1-559}$ and PknF$^{1-292}$, PknG$^{1-759}$ or PknB$^{1-279}$ as outlined in section 2.5.3.1. Purified PknG$^{1-759}$ and PknB$^{1-279}$ were a gift from the Division of Protein Structure (NIMR). The different proteins were incubated together with 200 µCi/ml [γ-^{32}P] ATP for 30 minutes, subjected to gel electrophoresis (Figure 5.9a) and radioactive bands were revealed by autoradiography (Figure 5.9b and c). Approximately 1 µg of protein was loaded into each lane. Figure a is a Coomassie blue stained gel and figure b is the result of autoradiography of the gel in figure a.

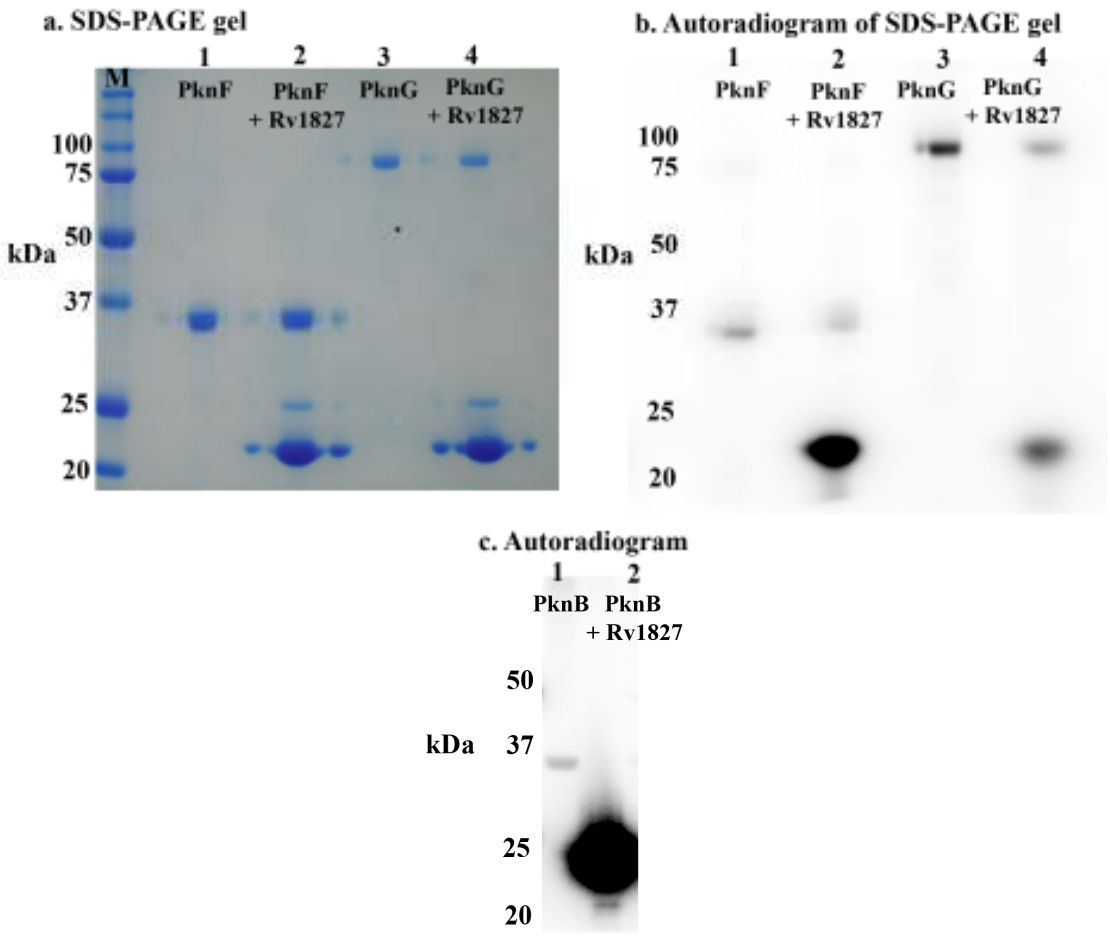


Figure 5.10 Phosphorylation of Rv1827 by multiple kinases. *In vitro* phosphorylation assays were performed with Rv1827$^{1-162}$ and PknF$^{1-292}$, PknG$^{1-759}$ or PknB$^{1-279}$ as outlined in section 2.5.3.1. The proteins were incubated together with 200 µCi/ml [γ-^{32}P] ATP for 30 minutes, subjected to gel electrophoresis (Figure 5.10a) and radioactive bands were revealed by autoradiography (Figures 5.10b and c). Approximately 1 µg of protein was loaded into each lane. Figure a is a Coomassie blue stained gel and figure b is the result of autoradiography of the gel in figure a. Rv1827$^{1-162}$ = 17 kDa, PknF$^{1-292}$ = 32 kDa, PknG$^{1-759}$ = 81 kDa, PknB$^{1-279}$ = 30 kDa.

shown in figure 5.10 and demonstrate that Rv1827$^{1-162}$ can indeed be phosphorylated *in vitro* by PknG (figure 5.10a and b lane 4) and PknB (figure 5.10c lane 2) demonstrating that the reason both these kinases did not phosphorylate Rv1747 was not because the kinases were non-functional.

Interestingly results in figure 5.10a and b lane 2 demonstrate that PknF can also phosphorylate Rv1827 showing that FHA domain-containing proteins can be phosphorylated by more than one serine-threonine protein kinase *in vitro*.

5.10 Discussion

The aim of this study was to explore the phosphorylation dependent interaction between PknF and Rv1747, to identify the sites on which PknF phosphorylated Rv1747 and to characterise the significance of these sites in terms of the phosphorylation status of Rv1747.

5.10.1 Protein purification

PknF$^{1-292}$ and Rv1747$^{1-559}$ were successfully expressed and purified from *E. coli* under non-denaturing conditions. Each protein's native conformation was retained throughout the complete purification; this means that the recombinant proteins were more likely to be folded into their correct conformation than if the proteins were purified under denaturing conditions and then refolded. However, it must be taken into consideration that the recombinant proteins consisted of only certain domains of the full proteins; the kinase domain of PknF was purified (residues 1 to 292 out of a total of 476) and the nucleotide binding and FHA domains of Rv1747 were purified (residues 1 to 559 out of 865). In the *M. tuberculosis* cell both these proteins are anchored in the cell membrane as indicated by the presence of transmembrane domains in both of the proteins. Therefore as the transmembrane domains were not included in the recombinant proteins it could be argued that it may have affected the folding of the proteins resulting in different conformations compared to when fixed in the cell membrane.

An interesting feature of recombinant PknF was that it separated on an SDS-PAGE gel as a doublet (figure 5.2). This is a feature quite common to recombinant protein kinases. For example, recombinant PknB also runs as a doublet (as shown in this study and Villarino *et al.*, 2005) as does PknG (figures 5.9 and 5.10). A reason why this may occur is because PknF appears to contain multiple autophosphorylation sites. This is further supported by the electrospray mass spectrometry data in figure 5.3 which highlighted the fact that $PknF^{1-292}$ can become autophosphorylated at multiple sites. It can be hypothesised that one of the autophosphorylation sites may result in a conformational change in the protein structure resulting in the kinase separating into two distinct protein bands.

Although it was possible to purify $PknF^{1-292}$ to homogeneity the purification of $Rv1747^{1-559}$ proved much more problematic. Initial solubility and purification trials proved unsuccessful; when very small amounts of soluble protein were obtained it seemed to bind irreversibly to the glutathione-S-transferase resin. This problem was overcome by expressing $Rv1747^{1-559}$ in BL21 star pRep4-*groESL* expression cells and through the addition of ten percent glycerol to all the protein purification buffers. Glycerol is a viscous liquid that is known to assist to keep proteins stable and furthermore it makes proteins less sensitive to freezing and thawing by helping to prevent protein damage caused by ice crystal formation. pRep4-*groESL* encodes GroEL and GroES, two bacterial molecular chaperones (Amrein *et al.*, 1995). Chaperones function to fold proteins correctly into their native conformation and thus improve the solubility of recombinant proteins by producing correctly folded and active proteins. GroEL is an Hsp60 family member and together with GroES they form a complex which hydrolyses ATP to provide energy for protein folding. The chaperone complex functions by repeatedly binding unfolded proteins, re-folding and then releasing proteins until they are folded into their native conformation (Horwich *et al.*, 2006). GroEL is approximately 60 kDa in size and GroES is 10 kDa in size. Purification of $Rv1747^{1-559}$ was achieved in the BL21 pRep4-*groESL* cells; $Rv1747^{1-559}$ was purified to near homogeneity but there was a low abundance of *E. coli* heat shock protein 70 (Hsp70) in all purifications. This protein was identified by MALDI-TOF mass spectrometry. The fact that the Hsp70

chaperone co-purified with Rv1747 suggests that this heat shock protein was required to maintain the solubility of the ABC transporter.

When aiming to conduct studies on the relationship between two proteins it is always optimal to purify both proteins to homogeneity. This is because the aim of the investigation is to observe the effects of the interaction between two proteins and not to obtain results as a consequence of contaminating proteins which inevitably leads to spurious results. However, for $Rv1747^{1-559}$ the best purification that could be achieved included a low abundance of Hsp70.

5.10.2 Autophosphorylation activity of recombinant PknF

Using an *in vitro* kinase assay recombinant $PknF^{1-292}$ was shown to be able to incorporate radio-labelled ATP and **was therefore demonstrated to be an active kinase capable of autophosphorylation**. Electrospray mass spectrometry analysis confirmed the presence of additional phosphate groups on the recombinant kinase after purification from *E. coli* ranging from plus two phosphate groups to plus seven indicating that *in vitro* $PknF^{1-292}$ is capable of autophosphorylation at numerous sites (figure 5.3).

Autophosphorylation by STPKs occurs in the activation loop of the kinase domain on specific threonine residues and autophosphorylation can occur more than once. The activation loop is situated in the catalytic domain of the kinase (see figure 1.4 for a schematic of a typical protein kinase showing the activation loop) (Villarino *et al.*, 2005). Autophosphorylation of PknF could occur in cis or trans, i.e. the phosphate group from one molecule could be transferred to its own activation loop or to an adjacent molecule's activation loop, both resulting in activation of the kinase (Lochhead, 2009). The solved structures of PknB, PknD and PknE kinase domains have shown them to be dimeric, which in the cell would presumably occur upon extracellular ligand binding to the sensor domain, and this dimerisation seemingly results in trans or cis autophosphorylation of the activation loop causing a conformational change leading to activation of the kinase (Gay *et al.*, 2006; Greenstein *et al.*, 2007; Mieczkowski *et al.*, 2008; Ortiz-Lombardia *et al.*, 2003; Scherr *et al.*, 2007; Young *et al.*, 2003). This

mechanism of activation is suggested to be conserved among other mycobacterial STPKs (Young *et al.*, 2003). For PknF, it is unknown whether a similar dimerisation event occurs; the predicted extracellular domain contains regions of low complexity and lacks a ligand binding domain with a recognisable function (Greenstein *et al.*, 2007).

Autophosphorylation can also occur in the juxtamembrane region between the kinase and transmembrane domains and also at the N-terminus of the protein. This was demonstrated for PknB, PknD, PknE and PknF in a study by Duran *et al* (Duran *et al.*, 2005). They showed using MALDI-TOF mass spectrometry combined with reverse-phase HPLC that PknF contained nine phosphorylated residues compared with PknB which contained five and PknD and PknE which both contained 11 (Duran *et al.*, 2005). Their PknF data supports the electrospray mass spectrometry data obtained in this study which showed that PknF contained up to seven additional phosphate groups but did not detect the remaining two additional phosphate groups found in the Duran *et al*. study.

In the 2005 study two phosphorylated activation loop threonine residues were identified for PknF, namely T173 and T175 (Duran *et al.*, 2005). For PknB, the same threonine residues were shown to be phosphorylated both in the Duran study and in previous investigations (Boitel *et al.*, 2003; Duran *et al.*, 2005; Young *et al.*, 2003). This reflects a degree of conservation between the activation mechanisms of the enzymes. Furthermore Duran and colleagues mutated the essential catalytic lysine residue to alanine in the PknB active site and showed that this mutant was not phosphorylated demonstrating that the phosphorylation observed in the experiments was the result of autocatalytic activity and not due to phosphorylation by an *E. coli* protein during the purification process (Duran *et al.*, 2005). Next, they focussed further on PknB and investigated whether phosphorylation of the activation loop affected kinase activity. They mutated the two phosphorylated active site threonine residues (T173 and T175) to alanines. Upon mutation of both these residues the kinase had 300 fold less activity compared to the wild type protein when assessed for its ability to phosphorylate myelin basic protein (Duran *et al.*, 2005). This result demonstrated that phosphorylation of the activation loop plays a significant role in the enzyme activity which has also been shown for another bacterial STPK in *Bacillus subtilis* (Madec *et al.*, 2003). Apart from the activation loop

threonines a further two phosphorylated residues were identified for PknF in the juxtamembrane region, namely threonine residue 287 and serine 290 but mutation of these residues did not affect kinase activity (Duran *et al.*, 2005) suggesting that they perform a different function in the signalling cascade.

5.10.2.1 The essential role of lysine 41 in PknF

As mentioned above, serine-threonine protein kinases have an essential catalytic lysine residue in their active site which is critical for enzyme activation due to its role in binding the α and ß phosphates of ATP (Kornev *et al.*, 2006; Young *et al.*, 2003). In this study it was examined whether the homologous lysine residue in PknF was also critical for phosphorylation. This was investigated to check that the autophosphorylation activity of $PknF^{1-292}$ observed in the *in vitro* phosphorylation experiment was the result of autocatalytic activity and not due to phosphorylation by an *E. coli* protein during the purification process. The lysine residue at position 41 of PknF was mutated to alanine and it was shown that this rendered the kinase inactive which was demonstrated by the fact that this strain did not incorporate radio-labelled ATP (figure 5.4). **This result showed the autophosphorylation activity of $PknF^{1-292}$ observed in the *in vitro* phosphorylation experiment was indeed the result of autocatalytic activity**. It has been previously shown that PknF K41M cannot phosphorylate myelin basic protein (MBP) (Koul *et al.*, 2001).

Interestingly, *E. coli* harbouring the $PknF^{1-292}$ K41A plasmid had the expected doubling time of 20 minutes up until induction of gene expression at an OD_{600} of 0.6; however the $PknF^{1-292}$ wild type strain had a much slower doubling time of approximately 35 to 40 minutes. Furthermore, when harvesting the cells after 18 hours of induction the *E. coli* cell pellets were also smaller in the wild type strain compared to the mutant strain suggesting that the doubling time of the *E. coli* strain expressing the wild type protein continued to be slower throughout the course of protein expression (data not shown). Control of recombinant protein expression in *E. coli* is rarely perfect i.e. in the absence of inducer, which in this system was IPTG, some protein expression often still occurs. If this was the case for the PknF expression constructs then it could explain why the doubling time of *E. coli* harbouring the $PknF^{1-292}$ wild type plasmid was considerably

slower: the active kinase was probably using ATP generated by *E. coli* to autophosphorylate. This in turn would result in there being less energy available for the *E. coli* cells to replicate resulting in the doubling time of this strain being considerably longer. In the PknF$^{1-292}$ K41A strain this was not the case as the mutation rendered the kinase incapable of binding ATP.

5.10.3 Rv1747 and PknF and other kinase/substrate interactions

In vitro phosphorylation assays confirmed that Rv1747 was a substrate for PknF (figure 5.5). This result verified the findings that Molle *et al*. published in 2004 but the present study further confirmed that the sites of phosphorylation must be within the Rv1747 protein itself as the GST-tag was cleaved off the recombinant protein leaving only five uncleaved amino acids at the N-terminus (GPLGS) (Molle *et al.*, 2004). However, it has to be noted that of the five amino acids one was a serine residue and thus had the potential of becoming phosphorylated by PknF.

Apart from the 2004 study investigating the interaction between PknF and Rv1747 (Molle *et al.*, 2004), numerous studies have been conducted which have characterised the interactions between other *M. tuberculosis* STPKs and their putative FHA domain containing substrates. A 2005 study by Villarino *et al*. used *M. tuberculosis* whole cell protein extracts, representing all the soluble proteins in *M. tuberculosis*, and investigated the ability of PknB to phosphorylate proteins in the extract. They demonstrated by two dimensional PAGE that PknB phosphorylated Rv1827, an FHA domain containing protein involved in the TCA cycle, and identified threonine residue 22 as the sole phosphate acceptor (Villarino *et al.*, 2005). Furthermore, in 2008 O'Hare and colleagues showed that PknG also phosphorylated Rv1827. They incubated Rv1827 with PknG and ATP and using mass spectrometry analysis showed that PknG phosphorylated Rv1827 at threonine residue 21 (O'Hare *et al.*, 2008) whereas PknB phosphorylated Rv1827 at threonine residue 22 (Villarino *et al.*, 2005). Furthermore, the phosphorylation of Rv1827 by PknB or PknG abrogated binding of Rv1827 to three proteins which are all involved in α-ketoglutarate metabolism (England *et al.*, 2009; Nott *et al.*, 2009). In 2003 it was demonstrated that EmbR was a substrate for PknH (Molle *et al.*, 2003). EmbR is involved in the biosynthesis of arabinogalactan, an important component of the

mycobacterial cell wall. Although only a few *M. tuberculosis* STPKs and their substrates have been characterised to date, it is clear that the interaction between kinases and their FHA domain containing substrates is involved in the regulation of key cellular processes in *M. tuberculosis*, previously predicted from bioinformatical analysis (Pallen *et al*., 2002).

5.10.3.1 Identification and characterisation of Rv1747 phospho-threonine residues

The next aim of this study was to identify the threonine residues on Rv1747 that were phosphorylated by PknF which would then allow for an investigation into how phosphorylation of Rv1747 by this kinase was controlling the function of the ABC transporter. The catalytic subunit of STPKs function to transfer a γ-phosphate from ATP to the side chain hydroxyl group of a serine or threonine residue in the substrate protein. Substrate phosphorylation can result in a conformational change in the protein structure or can lead to the recruitment of a phospho-serine/phospho-threonine binding domain in order to transduce the signal resulting in activation or repression of protein function. Two phosphorylated threonine residues were identified for Rv1747, namely T150 and T208. As depicted in figure 5.6 residue T150 is situated approximately half way between the FHA-1 domain and the FHA-2 domain. Residue T208, although not in the FHA-2 core region is situated just upstream of the domain. FHA domains function to bind phospho-threonine (pT) epitopes.

A study published in 2010 used a mass spectrometry based approach to identify phosphorylated proteins in *M. tuberculosis* (Prisic *et al.*, 2010). Of the 301 phosphoproteins they identified, 60 % had threonine residues as the phosphoacceptor and 40 % had serine residues as the phosphoacceptor. Furthermore they also found that PknA, PknB, PknD and PknG were themselves phosphorylated *in vivo* suggesting that these four kinases are all active in the growth conditions tested. They then combined bioinformatic analysis of the identified *in vivo* phosphorylation sites with data from *in vitro* kinase assays to identify phosphorylation site motifs for PknA, PknB, PknD, PknE, PknF and PknH (Prisic *et al.*, 2010). The motif identified for PknF is shown in figure 5.11. The motif of the six investigated kinases all included a threonine residue as the phosphoacceptor and hydrophobic residues at the pT+3 and pT+5 positions.

Furthermore, they found that the phosphoacceptor threonine could not be substituted for a serine for any kinase. For PknF, the overrepresented amino acids at the pT+3 position included phenylalanine, threonine, isoleucine and methionine and at the pT+5 position isoleucine was found to be the predominant residue (figure 5.11).

The phosphorylated threonine residues on Rv1747 identified in this study both share some of the features of the PknF preferred phosphorylation motif lending further support to the hypothesis that the identified phosphoacceptor threonine residues are real and are likely to be significant in terms of Rv1747 protein function. Site T208 has a methionine residue at pT+3 and an isoleucine residue at pT+5 (figure 5.6), both of which appear in the PknF preferred phosphorylation site motif. Furthermore, the identified T150 phosphorylation site contains an isoleucine residue at the pT+3 position which also appears in the predicted consensus motif; however, it does not contain an isoleucine or other hydrophobic residue at position pT+5 although it does have an alanine residue at position pT+6 which appears as one of the residues in the PknF motif.

By generating a model PknB structure in complex with a substrate peptide Prisic and colleagues then went on to map those residues on the peptide that were likely to come into contact with the active site residues of the kinase and potentially form hydrogen bonds (Prisic *et al.*, 2010). They showed that the central part of the PknB activation loop was in close contact with the substrate residues around the pT site, from residues pT-1 to pT+3. It was suggested that any of these sites could form hydrogen bonds with the kinase active site. As the pT+3 and pT+5 residues are conserved in the kinase motifs it suggested they are important for phosphorylation to occur. Furthermore the pT+3 position has been demonstrated to be important in FHA domain recognition (Liang & Van Doren, 2008) suggesting that STPKs and FHA domains have co-evolved to recognise overlapping consensus motifs.

Interestingly, the Prisic study also identified two phosphopeptides on Rv1747 in the *M. tuberculosis* protein lysates, both of which were not identified in our investigation: these were S161 and S178. Phosphorylated S161 was identified under all conditions tested: exponential and stationary phase, NO stress, peroxide stress, hypoxia and when grown with acetate as a carbon source. The S178 containing phosphoprotein was only identified

when *M. tuberculosis* had been growth with acetate as a carbon source (Prisic *et al.*, 2010). However, as the phosphorylation site motif analysis for PknF showed that the phosphoacceptor is a threonine residue and furthermore because FHA domains are phospho-threonine binding molecules, then one may argue that Rv1747 phosphorylation on a serine residue suggests that the phosphoacceptor site is not important for interaction with FHA domains or for the transduction of the signal leading to activation or repression of the function of the ABC transporter.

5.10.3.1.1 Regulation of Rv1747 by phosphorylation

One hypothesis into a possible mechanism of action of the Rv1747 protein is as follows. Upon PknF autophosphorylation the enzyme is rendered active, the peptide binding pocket of an Rv1747 FHA domain may then bind to one of the activation loop threonines. Binding of the FHA domain would also deliver the substrate protein into the active site of the kinase enabling substrate phosphorylation. This possible mechanism is shown in figure 5.12 (Villarino *et al.*, 2005). After substrate release, the FHA domain may then recognise and bind the pT epitope on Rv1747 causing a conformational change within the protein structure which would either confer it into an active or inactive conformation for the transport of the substrate across the *M. tuberculosis* cell membrane. As the T208 site is very close to the FHA-2 domain it would be impossible for the FHA-2 domain to rotate in such a way that it could bind the pT at position T208. Therefore one hypothesis raised during this study was that perhaps T208 phosphorylation causes FHA-2 dimerisation. The Rv1747 transporter is translated as a half-transporter which is predicted to dimerise to form the functional transporter and is anchored in the cell membrane. The hypothesis is that the FHA-2 domain from one half of the protein could bind to the pT208 of the other half of the protein and vice versa. The conformational change induced by this dimerisation could affect the function of the transporter. This hypothesised mechanism of regulation will be further explored in the subsequent chapter of this study. Furthermore, this mechanism of regulation has been previously observed in the human Chk2 serine-threonine protein kinase which is involved in the DNA damage response (Li *et al.*, 2008). Chk2 contains an N-terminal FHA domain and a C-terminal kinase domain. Li and colleagues showed that when Chk2 is phosphorylated at

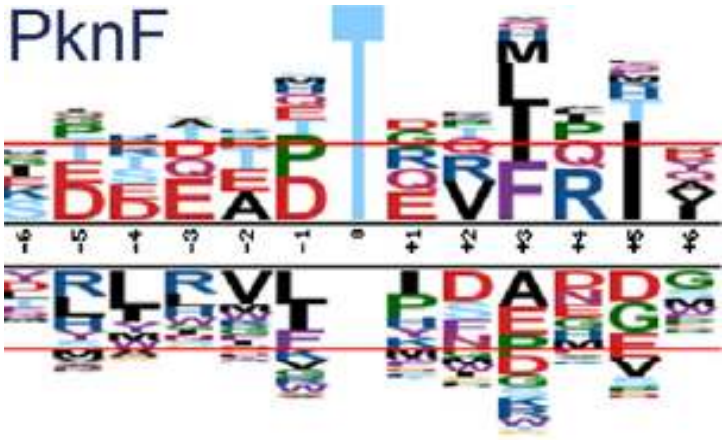


Figure 5.11 Phosphorylation site motif analysis for PknF. Figure adapted from Prisic *et al* 2010. At the 0 position is the pT phosphoacceptor site, residues above the midline were found to be overrepresented and those under the line were found to be underrepresented in their study. The red line indicates the 0.01 significance level. The figure shows that T is the preferred phosphoacceptor residue and hydrophobic residues were selected for at the pT+3 and pT+5 positions.

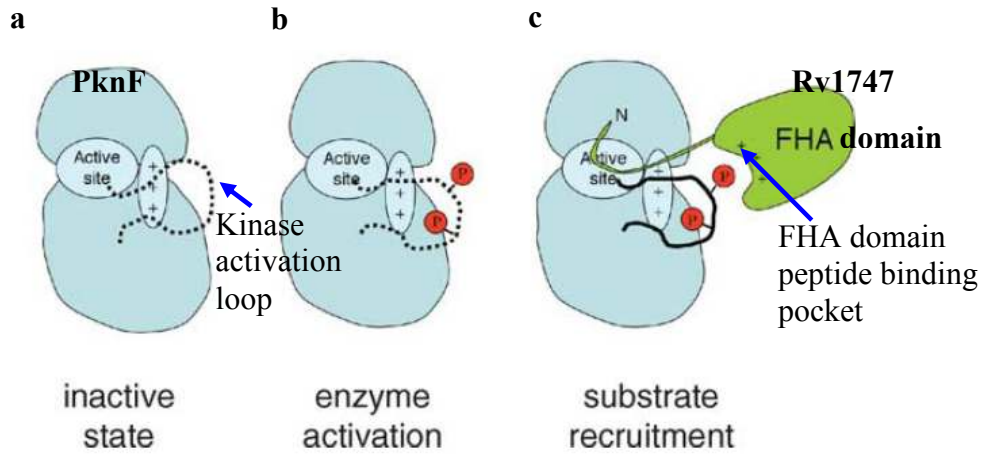


Figure 5.12 Proposed mechanism for the interactions between PknF and Rv1747. Figure adapted from Villarino *et al*. 2005. In the absence of autophosphorylation the kinase domain is in an inactive state (figure 5.12a). Upon autophosphorylation of the activation loop the kinase becomes enzymatically active (figure 5.12b). The Rv1747 FHA domains are then recruited to the kinase (figure 5.12c) where they bind an activation loop pT residue in their peptide binding pocket and then Rv1747 can be brought to the active site for phosphorylation on residues T150 and T208 (figure 5.12c).

position T68 the FHA domain of another molecule of Chk2 binds to the residue forming a tight dimer (Li *et al.*, 2008).

5.10.3.1.2 Mutation of Rv1747 residues T150 and T208

Mutation of both the Rv1747 T150 and T208 residues to alanines, rendering them incapable of receiving a donor phosphate group, dramatically reduced the level of Rv1747 phosphorylation observed in the *in vitro* phosphorylation assays (figure 5.8). **This result demonstrated that the two residues identified are both important for PknF phosphorylation of Rv1747**. Due to the fact that there was still a faint radio-labelled band in the double threonine to alanine mutant it cannot be ruled out that there is another threonine residue in Rv1747 that can also be phosphorylated (figure 5.8b lane 5). To investigate this LC-MS/MS could be repeated. In addition, when looking at the amount of radio-labelled ATP incorporated into both the Rv1747 single threonine mutant proteins it is clear that the T150A mutant incorporated less ATP compared with the T208A mutant, perhaps suggesting that the T150 site is more important in terms of protein phosphorylation and therefore protein function. Consequently, the results suggested that the T150 site is phosphorylated to a greater proportion compared to the T208 site although in the T208A mutant a reduction in incorporation of ATP can also be observed, but not to as great an extent as in the T150A mutant.

Interestingly, throughout all the *in vitro* phosphorylation assays with Rv1747 addition of various versions of Rv1747 to PknF resulted in a reduction or loss of the radio labelled PknF band (figures 5.5, 5.8 and 5.9). A similar occurrence was also noted in the Rv1827 experiments (figure 5.10). This loss of kinase activity is presumably through binding of an FHA domain to a pT residue in the kinase activation loop. At this stage one has to remember that after purification from *E. coli* PknF is already phosphorylated multiple times (figure 5.3c). Addition of radio labelled ATP only results in incorporation of radio labelled phosphate groups onto the residual threonine sites not already phosphorylated. In the case of PknF autophosphorylation, the incorporation of a radio labelled phosphate group could be in the activation loop or indeed on any other threonine residue on the protein which is accessible. Therefore the radio labelled kinase band may not be a result of incorporation of radio labelled phosphate groups into the activation loop. Furthermore

this means that FHA domain binding to an activation loop pT residue could occur in the absence of any incorporation of radio labelled phosphate groups. It is clear from these investigations that somehow binding of an FHA domain onto the PknF activation loop blocks further autophosphorylation activity of the kinase.

When the reactants are combined: PknF + Rv1747 + [γ-^{32}P]ATP an FHA domain would be able to bind to the activation loop of PknF immediately if a threonine in this loop was already phosphorylated (during purification PknF can use ATP generated by *E. coli* to autophosphorylate itself). A phosphate group from the addition of radio labelled ATP could then be used to phosphorylate the substrate protein. This could explain why we see a reduction or loss of kinase activity upon addition of substrate proteins.

Alternatively, perhaps when one of the FHA domains docks to a pT residue in the activation loop of PknF it does not dissociate until the substrate is phosphorylated in the active site. If the active site of the enzyme does not or can not phosphorylate a threonine residue at positions T150 or T208 of Rv1747, then in a sense the mutated Rv1747 protein may act as an inhibitor of kinase activity as the FHA domain would remain bound to the pT residue in the activation loop and therefore the radio-labelled kinase band on the autoradiogram could become diminished.

The hypothesised explanations discussed above may also offer an explanation for the results obtained when it was investigated whether two other STPKs could phosphorylate Rv1747. Results showed that when PknG and Rv1747 (figure 5.9b lane 4) or PknB and Rv1747 (figure 5.9c lane 2) were incubated together an inhibition of kinase activity was observed. Rv1747 was shown not to be a substrate for PknG and to be a very poor substrate for PknB. In this instance the threonine sites were not mutated but Rv1747 was not a substrate for these two kinases. Therefore one could hypothesise that an FHA domain could dock to a pT residue in the kinases activation loop and although the enzymes could not then phosphorylate Rv1747, the action of the FHA domain binding to the activation loop could have resulted in a reduction or loss of kinase activity.

Interestingly, a study by Prisic *et al*. found that PknG and PknK had a distinctly different phosphorylation consensus motif compared with PknF, PknA, PknB, PknD, PknE and

PknH perhaps offering an explanation as to why PknG could not phosphorylate Rv1747 at all *in vitro* (Prisic *et al.*, 2010). Furthermore the soluble proteins PknG and PknK contain different kinase domains compared with the other STPKs (Narayan *et al.*, 2007) suggesting why they would probably have different preferred phosphorylation motifs compared with the transmembrane bound kinases.

This study only assessed the ability of PknF, PknB and PknG to phosphorylate Rv1747; there are eight other STPKs in the *M. tuberculosis* genome that could also phosphorylate Rv1747 not only *in vitro* but also *in vivo*. For example, *pknE* is situated in relatively close proximity upstream of *Rv1747* in the genome so perhaps this kinase could play a role in controlling Rv1747 function; furthermore, the PknE preferred phosphorylation consensus motif is highly similar to that of PknF (Prisic *et al.*, 2010). In addition, a publication in 2005 demonstrated that PknE, PknD, PknB and PknF could all phosphorylate one or both Rv1747 FHA domain-containing constructs (Grundner *et al.*, 2005). PknE was shown to only phosphorylate the FHA-1 domain-containing construct and coupled with its close genomic localisation to *Rv1747* it therefore warrants future work investigating the possible role this kinase plays in controlling Rv1747 function. PknE may phosphorylate Rv1747 on different threonine residues compared with PknF as was demonstrated for Rv1827 with PknB and PknG (O'Hare *et al.*, 2008).

5.10.4 Rv1827 *in vitro* phosphorylation assays

To confirm that recombinant PknG and PknB were functional kinases it was assessed whether they could phosphorylate Rv1827. The results showed that $Rv1827^{1-162}$ can indeed be phosphorylated *in vitro* by PknG, PknB and PknF demonstrating that the reason PknB and PknG did not phosphorylate Rv1747 was not because the kinases were non-functional but was because Rv1747 was not a substrate for these kinases. Furthermore, the results demonstrated that FHA domain-containing proteins can be phosphorylated by more than one serine-threonine protein kinase *in vitro*. These data corroborate previous data showing that Rv1827 can be used as an *in vitro* substrate by PknF, PknD and PknE (Villarino *et al.*, 2005).

Moreover, it is clear from the results of this study that PknB is the preferred kinase for Rv1827 as the incorporated radio labelled signal is much greater when Rv1827 is phosphorylated with PknB compared to PknG. PknB and PknG phosphorylate Rv1827 on two distinct threonine residues. The importance of PknB phosphorylation on Rv1827 function was described in a publication by Nott *et al*. where they demonstrated that phosphorylation of a threonine residue by PknB at the N terminus of Rv1827 caused an intramolecular association with the Rv1827 FHA domain, blocking the association of this domain with its three binding partners which are all key metabolic enzymes involved in the α-ketoglutarate metabolism in the TCA cycle (Nott *et al.*, 2009). However, a proteomics based study only conclusively identified T21 (the residue phosphorylated by PknG) as a phosphoacceptor site in Rv1827 (Prisic *et al.*, 2010).

5.10.5 *In vitro* phosphorylation assays

The use of *in vitro* phosphorylation assays to study protein-protein interactions is a widely accepted method which has helped many scientists derive the function of specific proteins. However, it must be taken into consideration that all the assays in this study were performed *in vitro* and an *in vitro* substrate does not confirm an *in vivo* physiological substrate. In more general terms, inside the *M. tuberculosis* cell there are both temporal and spatial separations of protein expression; proteins can be expressed at different times throughout replication or throughout the course of infection and they can also be expressed at different places in the cell, both cases meaning that two proteins might never actually come into contact with one another. This is exemplified by the result from this study which showed that PknF could phosphorylate Rv1827 *in vitro* (figure 5.10). In the *M. tuberculosis* cell Rv1827 is involved in α-ketoglutarate metabolism in the TCA cycle; it therefore seems extremely unlikely that PknF would also have an involvement in the regulation of this protein therefore suggesting that Rv1827 is not a physiological *in vivo* substrate for PknF. This is the reason why one must be very careful when performing *in vitro* assays and it must be ensured that findings are always followed up with *in vivo* experiments.

In this present study the Rv1747 threonine mutants have been analysed in both *in vitro* and *in vivo* experiments. Thus in the next chapter murine bone marrow derived

macrophage infection and murine aerosol infection experiments have been performed to deduce the effect of Rv1747 phosphorylation within the *M. tuberculosis* cell by using phenotype as readout of protein function due to the fact that the *Rv1747* deletion strain is attenuated for growth in these models.

5.10.6 Summary

In this chapter it has been shown that **PknF does indeed play a role in a signal transduction system with Rv1747 and using mass spectrometry two phosphorylated threonine residues on the ABC transporter have been identified which were shown to be critical for the phosphorylation of Rv1747 by PknF**. Furthermore, this data also gives credence to the hypothesis that PknF controls Rv1747 function by phosphorylation thus warranting further investigation in the final chapter of the study where the significance of the phosphorylated threonine sites will be determined by both *in vitro* and *in vivo* experiments. In addition, the importance of the FHA domains will be characterised and the question as to whether phosphorylation by PknF positively or negatively regulates the activity of the Rv1747 ABC transporter will be explored.

CHAPTER 6. HOW IS PknF CONTROLLING Rv1747 FUNCTION? PART 2

FURTHER CHARACTERISATION OF THE *pknF-Rv1747* SIGNALLING SYSTEM AND ANALYSIS OF THE IMPORTANCE OF BOTH GENES FOR THE VIRULENCE OF *M. tuberculosis*

6.1 Introduction

Upon inhalation of *M. tuberculosis* bacilli, the bacteria enter the respiratory tract of the host and are phagocytosed by alveolar macrophages lining the lungs. *M. tuberculosis* has the ability to evade the host's defence mechanisms by inhibiting phagosome-lysosome fusion allowing survival and replication within the macrophages until the infection usually becomes contained within a granulomatous lesion. Macrophages derived from the bone marrow of mice (BMDMs) are used in many areas of *M. tuberculosis* research which include investigating the effects of *M. tuberculosis* gene disruption. *M. tuberculosis* typically grows and replicates inside BMDMs, just like in human alveolar macrophages, and they are therefore used as a model for experiments assessing the growth and survival of different *M. tuberculosis* strains. In this study the growth and survival of the *ΔRv1747*, *ΔpknF* and various versions of the *Rv1747* complemented strains containing point mutations in key domains or in phosphorylation sites were assessed.

Previous investigations in this study (Chapter 5) had determined that PknF does indeed play a role in a signal transduction system with Rv1747 and using mass spectrometry two phosphorylated threonine residues on the ABC transporter were identified which were shown to be critical for the phosphorylation of Rv1747 by PknF. These findings gave further credence to the hypothesis that PknF controls Rv1747 function by phosphorylation. The aim of this final chapter was to determine the significance of the phosphorylated threonine sites, to characterise the importance of the FHA domains in terms of protein regulation, to investigate whether phosphorylation by PknF positively or negatively regulates the activity of the Rv1747 ABC transporter and to propose a

model of how PknF controls Rv1747 based on the data generated in this study as a whole.

6.2 Assessment of the *Rv1747* null mutant in a murine macrophage infection model

The first aim of this study was to determine whether the *Rv1747* gene plays a role in the virulence of *M. tuberculosis*. The hypothesis was that the *Rv1747* deletion strain would be attenuated for growth in both naïve and IFNγ activated macrophages; a previous investigation in 2005 had discovered that the *ΔRv1747* strain was attenuated for growth in naïve macrophages, dendritic cells and in the lungs and spleens of mice (Curry *et al.*, 2005). To further explore this hypothesis, the growth of wild type H37Rv, *ΔRv1747* and *Rv1747* complement strains were assessed in naïve and IFNγ activated murine bone marrow derived macrophages (BMDMs) over an infection time course of 168 hours. Macrophages were derived from six to eight week old female BALB/c mice as described in section 2.4.12.1, activated as required with 10 ng/ml IFNγ and infection with *M. tuberculosis* strains was performed as outlined in section 2.4.12.2. At 6, 24, 72, 120 and 168 hours post infection macrophages were lysed and *M. tuberculosis* surviving within the macrophages were enumerated. Growth of *M. tuberculosis* colonies on 7H11 agar plates was monitored for at least four weeks to ensure that all colonies were counted and to correct for any differences in the time colonies took to grow between the strains.

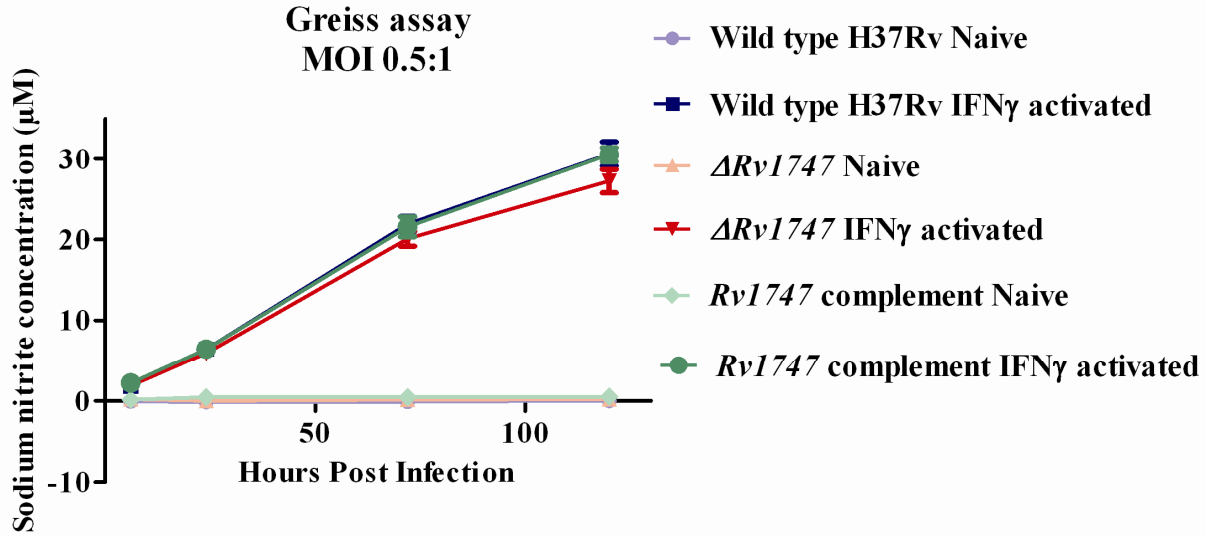
To confirm that treatment with IFNγ activated the BMDMs a Greiss nitrite assay was performed. Nitrite (NO_2^-) is a stable breakdown product of nitric oxide and thus can be used as an indicator of macrophage activation using the Greiss assay. Naïve and activated BMDMs were infected with wild type H37Rv, *ΔRv1747* and *Rv1747* complementing strains and the concentration of NO/nitrite present in the supernatants of the infected BMDMs was assessed over 120 hours of infection as described in section 2.4.12.3. Naïve BMDMs produced virtually no measurable concentration of NO/nitrite over the course of the experiment when infected with either of the three *M. tuberculosis* strains (figure 6.1). However, BMDMs pre-stimulated with IFNγ prior to infection and then replaced with media containing IFNγ after removal of the extracellular *M.*

tuberculosis following six hours of infection did indeed produce NO/nitrite showing that the activation procedure had been successful (figure 6.1). Results also demonstrated that there were no differences observed in the levels of NO/nitrite produced when infected with the three different *M. tuberculosis* strains. Sodium nitrite concentrations ranged from an average of 1.9 μM at 6 hours to 29.5 μM at 120 hours post infection in IFNγ stimulated cells. The results presented verified that BMDMs are indeed activated when pre-incubated with IFNγ for 18 hours prior to infection with *M. tuberculosis* and maintain an increasing concentration of NO/nitrite throughout the course of infection.

Results from a BMDM infection experiment in both naïve (figure 6.2a) and IFNγ activated cells (figure 6.2b) showed that in naïve cells the bacterial load of the wild type and *Rv1747* complement strains of *M. tuberculosis* continued to increase throughout the experiment. However, in the *Rv1747* deletion strain, although a similar number of bacilli were phagocytosed by the macrophages as indicated by the data at the 6 hour uptake time point, this strain replicated significantly slower and by 168 hours post infection the bacterial load of this strain was approximately 10 fold lower compared to that of the wild type and complementing strains. This growth attenuation was shown to be statistically significant at 168 hours post infection ($p<0.01$).

The higher NO/nitrite concentration within the activated BMDMs were seen to limit the *M. tuberculosis* replication at later time points (figure 6.2b). However, the wild type and *Rv1747* complementing strains again demonstrated a similar pattern of growth to one another whilst the *ΔRv1747* strain appeared to grow very little and the bacterial load of this strain actually decreased between 120 and 168 hours post infection. This growth attenuation was shown to be statistically significant at 120 ($p<0.01$) and 168 ($p<0.001$) hours post infection.

Taken together, the results obtained confirm that an *Rv1747* deletion strain is attenuated for growth and supports the hypothesis that Rv1747 is important for the growth and replication of *M. tuberculosis* in BMDMs. These results corroborate an earlier study where the *Rv1747* deletion strain was shown to be attenuated for growth in both BMDMs and mice (Curry *et al.*, 2005).



Hours Post Infection	Wild type H37Rv Naive			Wild type H37Rv IFNγ activated			ΔRv1747 naive			ΔRv1747 IFNγ activated			Rv1747 complement naive			Rv1747 complement IFNγ		
6	0.01	0.01	-0.19	1.68	2.00	1.79	0.64	-0.09	0.11	1.68	2.00	1.89	0.22	0.11	0.11	2.73	1.79	2.10
24	-0.09	-0.09	-0.19	6.08	6.60	6.60	-0.09	-0.09	0.11	5.55	5.97	6.18	0.64	0.32	0.32	6.29	6.08	6.91
72	0.11	-0.19	-0.19	21.88	20.94	22.82	0.32	0.01	0.11	19.16	20.10	21.04	0.43	0.53	0.32	20.52	21.25	22.93
120	0.22	0.01	-0.19	29.52	29.94	32.34	0.01	0.11	0.53	26.48	26.27	28.89	0.53	0.53	0.43	30.36	29.73	31.51

Figure 6.1 Greiss nitrite assay after infection of BMDMs with wild type H37Rv, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. The concentration of NO/nitrite present in the supernatants of infected BMDMs was assayed over 120 hours as described in section 2.4.12.3. The concentration of NO/nitrite was calculated from naïve and IFNγ activated BMDMs infected with wild type H37Rv, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. Data plotted are the mean ± S.D of three technical replicates assayed in duplicate. There is a significant difference between the concentration of sodium nitrite in naïve and IFNγ activated cells from 24 hours onwards.

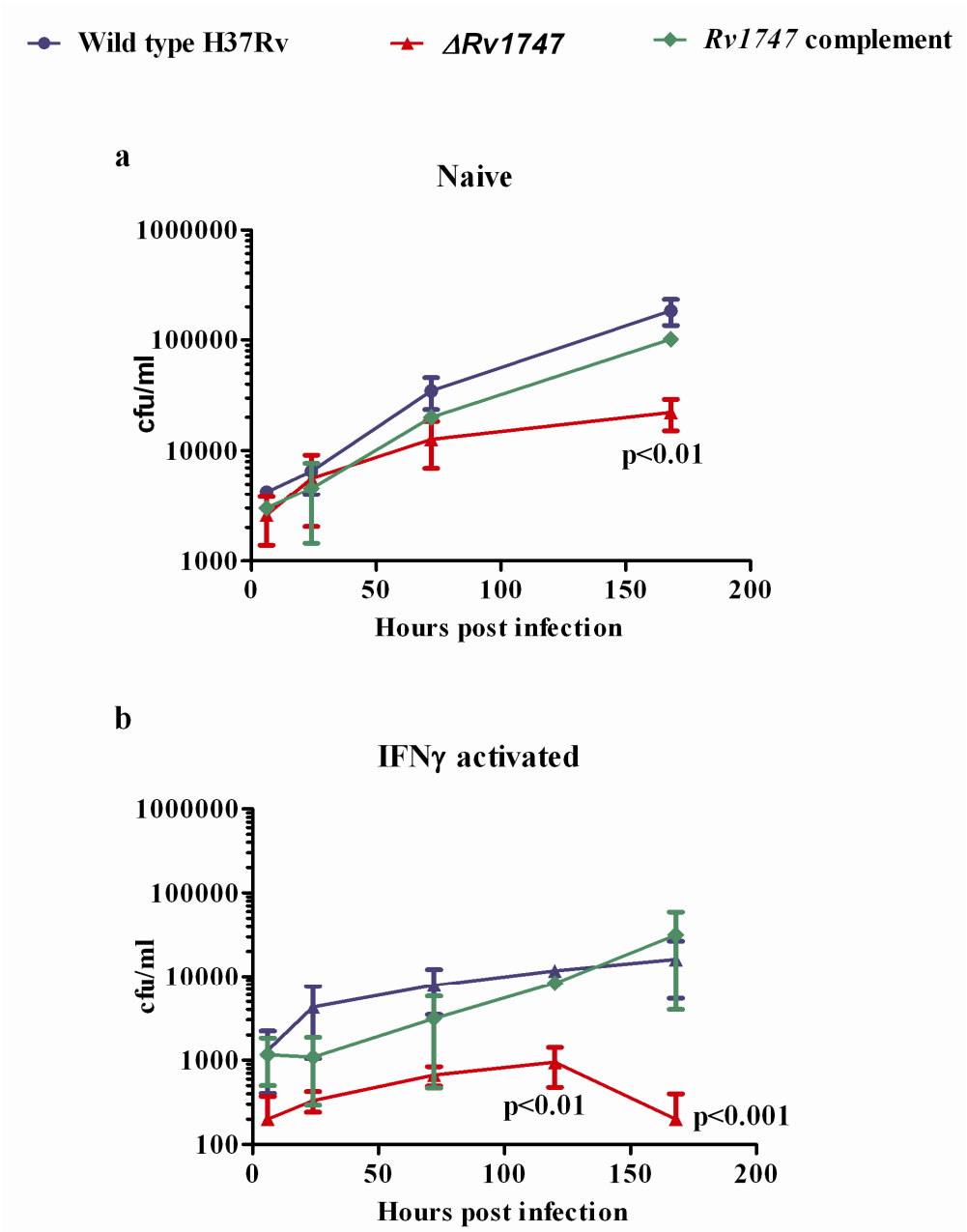


Figure 6.2 Bone marrow derived macrophage infection with wild type H37Rv, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. Infection of BMDMs with *M. tuberculosis* strains was performed with an MOI of 0.5:1 (*M.tb*:macrophage) as outlined in section 2.4.12.2. Results show the growth of wild type (blue), *ΔRv1747* (red) and *Rv1747* complement (green) strains over 168 hours. Figure a shows the growth of *M. tuberculosis* strains in naïve macrophages and figure b shows the growth of *M. tuberculosis* strains in IFNγ activated cells. Error bars indicate the mean ±S.D of three technical replicates. Data show that the *ΔRv1747* strain is attenuated for growth in both naïve and IFNγ activated macrophages. p-values are derived from an unpaired Students t-test between the wild type and *ΔRv1747* strains.

6.3 Assessment of the growth of the *Rv1747* phosphorylation-mimic and phosphorylation-null mutants in macrophage and murine infection models

Previous investigations in this study had identified two threonine residues on Rv1747 that were phosphorylated by PknF (figure 5.6); furthermore both these pT sites were demonstrated to be important for the *in vitro* phosphorylation of Rv1747 by PknF (figure 5.8). The working hypothesis was that PknF controls the function of the ABC transporter in a phosphorylation dependent manner. It was therefore hypothesised that both the identified phospho-threonine sites on Rv1747 may be important for Rv1747 transporter function. In view of the fact that the *M. tuberculosis ΔRv1747* strain was shown to be attenuated for growth in both naïve and IFNγ activated BMDMs (figure 6.2) this same macrophage infection model was employed to assess whether the two identified pT sites were important for Rv1747 function using phenotype as a readout of protein function. The aim was to mutate the T150 and T208 residues within the *Rv1747* complementing plasmid and then transform these plasmids into the *ΔRv1747* strain and assess each strain for its ability to complement the growth phenotype observed in the *ΔRv1747* mutant strain in BMDMs.

6.3.1 Construction of the panel of threonine mutations

As it was not known whether phosphorylation of Rv1747 by PknF positively or negatively regulated the activity of the ABC transporter two types of threonine mutations were created in the *Rv1747* complementing plasmid. The first mutation altered the threonine residues to alanine residues rendering each site incapable of phosphorylation (phospho-null); the second mutation changed the threonine residues to aspartic acid. Aspartic acid is an amino acid whose structure mimics the structure of a phosphorylated threonine (pT) residue and therefore these mutants should act as constitutive phosphorylation mimics (phospho-mimic). Site directed mutagenesis was performed as described in section 2.2.3 and a list of primers can be found in appendix II. Six mutations were made in all, namely T150A, T208A, T150A/T208A, T150D, T208D and T150D/T208D. The panel of *Rv1747* complementing plasmids were each transformed into the *ΔRv1747* strain and the BMDM infection experiments were then

performed to assess each strains' ability to complement the growth phenotype observed in the mutant strain. It was hypothesised that if phosphorylation by PknF inhibited the transporter activity of Rv1747 then we may see an attenuation of growth in the *Rv1747* T150D/T208D complemented strain whereas if phosphorylation activated the transporter activity of Rv1747 then it was hypothesised that a growth phenotype might be observed in the *Rv1747* T150A/T208A complementing strain.

6.3.2 Assessment of the growth of the *Rv1747* phosphorylation-mimic and phosphorylation-null double mutants in a macrophage infection model

The growth of the *Rv1747* complement T150A/T208A and *Rv1747* complement T150D/T208D strains were assessed in a BMDM infection experiment. It was hypothesised that if one of the mutations rendered Rv1747 non-functional then the *ΔRv1747* growth phenotype would be also observed with this strain. Figure 6.3 shows the results from a BMDM infection experiment in both naïve (figure 6.3a) and IFNγ activated cells (figure 6.3b). Figure 6.3 shows under both conditions the *ΔRv1747* strain is attenuated for growth compared with the wild type and *Rv1747* complement strains. In the naïve BMDM infection the growth of the *Rv1747* complement T150A/T208A and the *Rv1747* complement T150D/T208D strains were significantly different from the wild type strain at 120 and 168 hours post infection ($p<0.05$). In the IFNγ activated macrophages both the *Rv1747* complement T150A/T208A and the *Rv1747* complement T150D/T208D strains had an intermediate growth phenotype between that of the wild type and the *ΔRv1747* strains, the growth difference were shown to be statistically significant from the wild type strain at 168 hours post infection ($p<0.05$).

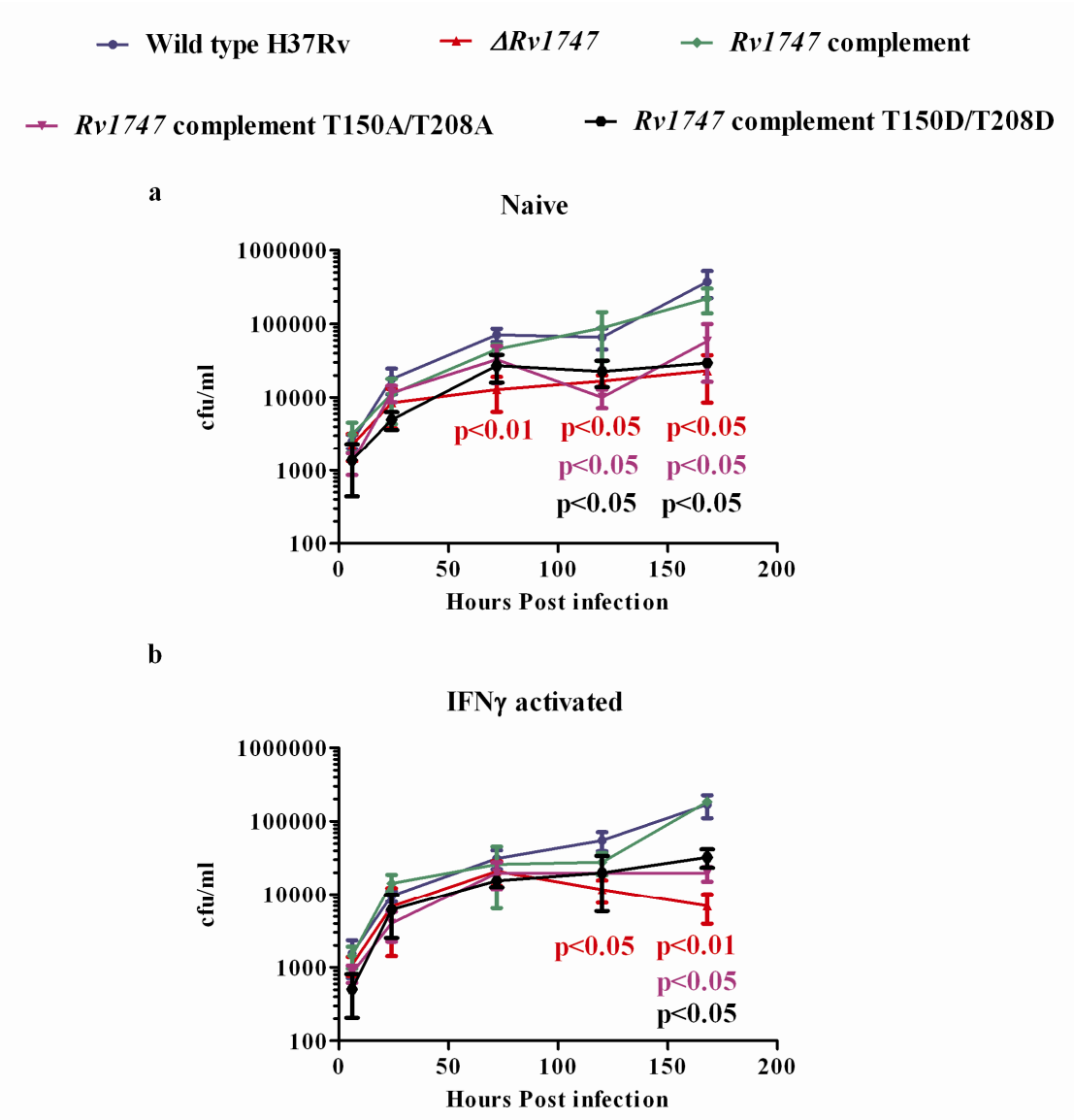


Figure 6.3 Bone marrow derived macrophage infection with wild type H37Rv, *ΔRv1747*, *Rv1747* complement, *Rv1747* complement T150A/T208A and *Rv1747* complement T150D/T208D *M. tuberculosis* strains. Results show the growth of the five strains over 168 hours. Figure a shows the growth of *M. tuberculosis* strains in naïve macrophages and figure b shows the growth of *M. tuberculosis* strains in IFNγ activated cells. Data show a representative experiment from two biological replicates performed technical triplicate. Error bars indicate the mean ±S.D of three technical replicates. Data show that the *ΔRv1747* strain is attenuated for growth in both naïve and activated macrophages. p-values are derived from unpaired Students t-tests between the wild type and *ΔRv1747*, *Rv1747* complement T150A/T208A, or *Rv1747* complement T150D/T208D strains.

6.3.3 Assessment of the growth of the *Rv1747* phospho-mimic and phospho-null double mutants in a murine aerosol infection model

The next aim of this study was to determine whether the *Rv1747* complement T150A/T208A and *Rv1747* complement T150D/T208D strains had a growth phenotype in a murine aerosol model of infection. *M. tuberculosis* strains were prepared for infection into female BALB/c mice as described in section 2.4.13. *M. tuberculosis* aerosol infections were performed by Angela Rodgers and Barry Walker at the National Institute for Biological Standards and Control. Results are shown in figure 6.4. Each time point consisted of five mice and the viable *M. tuberculosis* colonies were enumerated from both the lungs and spleens. Data from the lungs shows that the *ΔRv1747* strain is attenuated for growth compared with the wild type strain (figure 6.4a). The *Rv1747* full complement strain grew similarly to the wild type up until day 90 when the growth of this strain and the two mutated complement strains decreased in CFU (figure 6.4a).

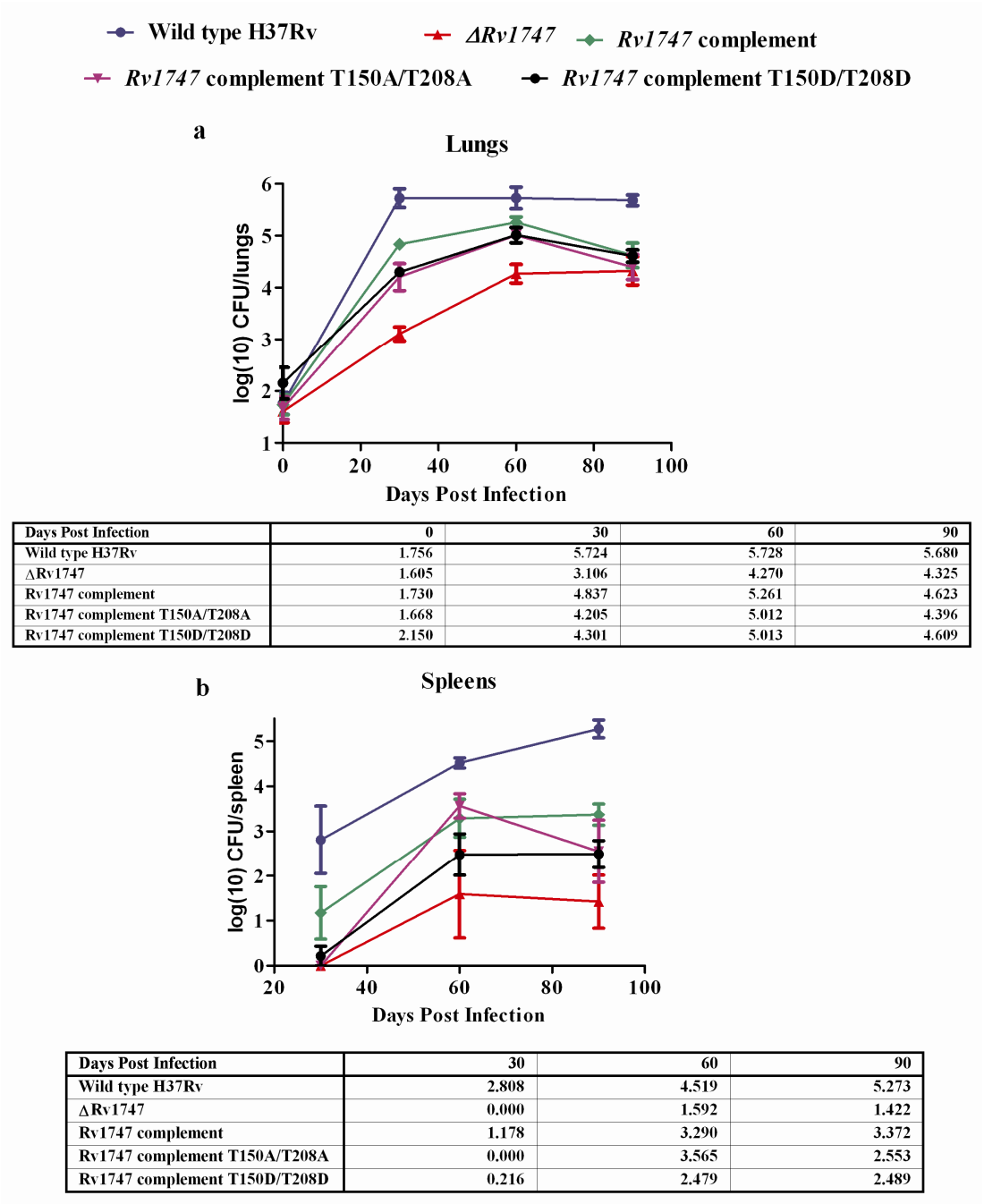
Interestingly, at day 30, mice infected with the *ΔRv1747* and *Rv1747* complement T150A/T208A strains had no *M. tuberculosis* bacilli present in the spleens (figure 6.4b). Furthermore, at day 60 only three of the five animals infected with the *ΔRv1747* strain had bacilli in the spleens suggesting that both this mutant and the *Rv1747* complement T150A/T208A strain are defective for growth in the spleens. Furthermore, the *Rv1747* complement T150D/T208D strain also did not grow as well as the *Rv1747* complement in the spleens. Overall, in the spleens the *Rv1747* complementing strain had an intermediate growth phenotype between that of the *ΔRv1747* and wild type strains.

6.3.4 Assessment of the growth of the *Rv1747* phospho-mimic and phospho-null single mutants in a murine macrophage infection model

It was assessed whether any of the single threonine to alanine or threonine to aspartic acid mutant constructs had a growth phenotype in BMDM infection experiments. Experiments were performed to determine whether either one of the identified pT residues were particularly important for the regulation of the function of the ABC transporter. Indeed, from the *in vitro* phosphorylation assays it was concluded that the

T150 site was probably more important than the T208 site due to the level of phosphorylation loss observed when the T150A mutation was introduced into the Rv1747 protein expression construct (figure 5.8). It was hypothesised that if one residue was particularly important then a growth phenotype may be observed in that mutant in BMDMs if it affected Rv1747 protein function.

The growth of the wild type, *ΔRv1747*, *Rv1747* complement, *Rv1747* complement T150A, *Rv1747* complement T150D, *Rv1747* complement T208A and *Rv1747* complement T208D *M. tuberculosis* strains in naïve BMDMs was assessed (figure 6.5). As previously observed, results showed that the *ΔRv1747* strain is attenuated for growth compared with the wild type and *Rv1747* complement strains at 120 and 168 hours post infection. The *Rv1747* complement T150A and *Rv1747* complement T208A strains followed a very similar growth pattern to that of the wild type and *Rv1747* complement strains throughout the time course of infection. There were no significant differences between the bacterial loads of the single mutant strains and the wild type strain apart from in *Rv1747* complement T150D strain at 120 hours post infection. At 168 hours post infection there was no longer a statistically significant difference between the growth of this strain and the wild type strain. At 120 hours post infection the *Rv1747* complement T208D strain had a calculated bacterial load which was in between that of the wild type and *Rv1747* deletion strain although this was not deemed to be a significant difference.



Days Post Infection	0	30	60	90
Wild type H37Rv	1.756	5.724	5.728	5.680
ΔRv1747	1.605	3.106	4.270	4.325
Rv1747 complement	1.730	4.837	5.261	4.623
Rv1747 complement T150A/T208A	1.668	4.205	5.012	4.396
Rv1747 complement T150D/T208D	2.150	4.301	5.013	4.609

Days Post Infection	30	60	90
Wild type H37Rv	2.808	4.519	5.273
ΔRv1747	0.000	1.592	1.422
Rv1747 complement	1.178	3.290	3.372
Rv1747 complement T150A/T208A	0.000	3.565	2.553
Rv1747 complement T150D/T208D	0.216	2.479	2.489

Figure 6.4 Murine aerosol infection experiment with wild type H37Rv, *ΔRv1747*, *Rv1747* complement, *Rv1747* complement T150A/T208A and *Rv1747* complement T150D/T208D M. tuberculosis strains. *M. tuberculosis* strains were prepared for infection into female BALB/c mice as described in section 2.4.13. The results for each time point are the means of the CFU determinations performed on organs from five mice and the error bars show the standard error of the means. The growth of the *ΔRv1747* strain is significantly different compared with the wild type in the lungs and spleens at 30, 60 and 90 days post infection ($p<0.05$). The growth of the *Rv1747* complement T150A/T208A and *Rv1747* complement T150D/T208D strains are significantly different compared with the wild type in the spleens at 30, 60 and 90 days post infection and in the lungs at 30 and 90 days post infection ($p<0.05$).

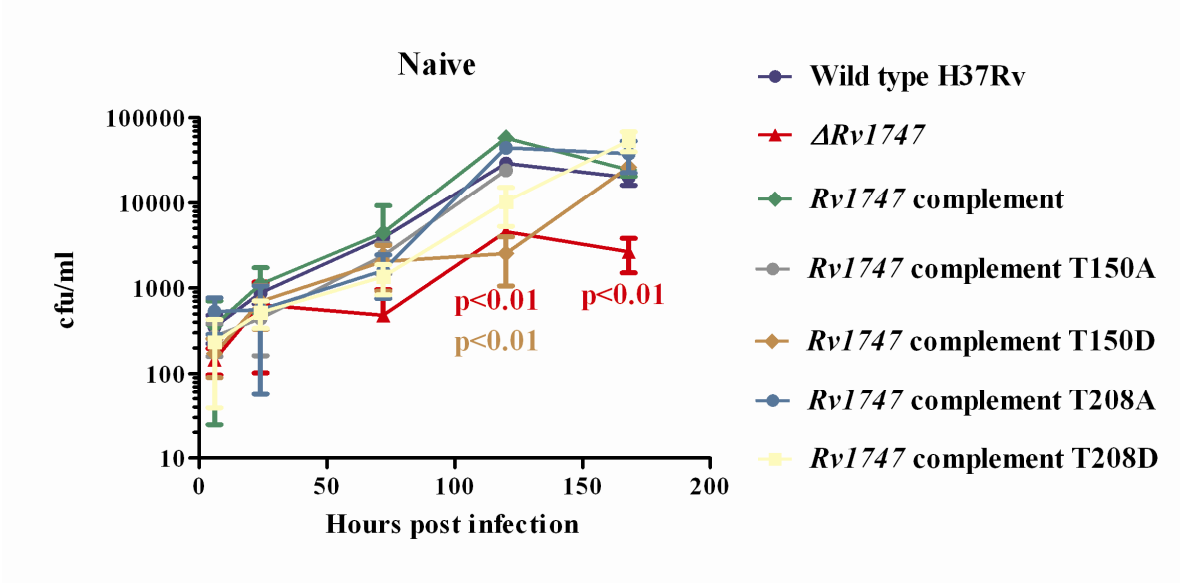


Figure 6.5 Bone marrow derived macrophage infection with wild type H37Rv, *ΔRv1747*, *Rv1747* complement, *Rv1747* complement T150A, *Rv1747* complement T150D, *Rv1747* complement T208A and *Rv1747* complement T208D *M. tuberculosis* strains. Results show the growth of wild type, *ΔRv1747*, *Rv1747* complement and the four single mutant strains in naïve BMDMs over 168 hours. Data show a representative experiment from two biological replicates performed technical triplicate. Error bars indicate the mean ±S.D of three technical replicates. Data show that the *ΔRv1747* strain is attenuated for growth in naïve macrophages. p-values are derived from unpaired Students t-tests between the wild type and *ΔRv1747* or *Rv1747* complement T150D strains.

6.4 Analysis of cytokine profiles

One hypothesis raised in this study was that perhaps the reason why the *ΔRv1747* strain demonstrated a growth attenuation phenotype in BMDMs was because infection with this strain induces a different cytokine response from the BMDMs which leads to inhibition of growth or increased killing of the *ΔRv1747* strain. A study examining the effects of the deletion of the *pknE* gene found that infection of THP-1 macrophages with this mutant caused the THP-1 cells to secrete significantly lower levels of the pro-inflammatory cytokines TNFα and IL6 in the culture supernatants compared to the wild type and complementing strains (Jayakumar *et al.*, 2008). It was therefore of interest to examine the cytokines secreted by BMDMs after infection with wild type H37Rv, *ΔRv1747* and *Rv1747* complementing strains. The secretion of three cytokines were examined throughout the course of infection; IL12-p40, TNFα and IL6. Each figure represents the results of one biological replicate performed in technical triplicate.

Figure 6.6 shows the results of IL12-p40 secretion after infection with all three strains at a multiplicity of infection (MOI) of 0.5:1 (*M.tb*:macrophage) (figure 6.6a) and 5:1 (figure 6.6b). An MOI of 0.5:1 was used for all BMDM infection experiments where the growth of *M. tuberculosis* was enumerated. As expected, infection with a higher MOI and a greater number of macrophages seeded per well resulted in a greater secretion of IL12-p40 and furthermore in both figures at each individual time point the amount of secreted cytokine produced by naïve cells was always significantly lower that that of the IFNγ activated cells. Figures 6.6a and 6.6b show that at both MOIs there were no differences observed in the levels of secreted cytokine produced by infection with the three *M. tuberculosis* strains.

Figure 6.7 shows the results of TNFα secretion; again, as predicted, infection with a higher MOI resulted in a higher concentration of TNFα to be secreted by the macrophages and furthermore in figure 6.7b at each individual time point the amount of secreted cytokine produced by naïve cells was always lower that that produced by the IFNγ activated cells. Figure 6.7b shows there were no significant differences in the concentrations of secreted TNFα produced between the wild type, *ΔRv1747* and *Rv1747*

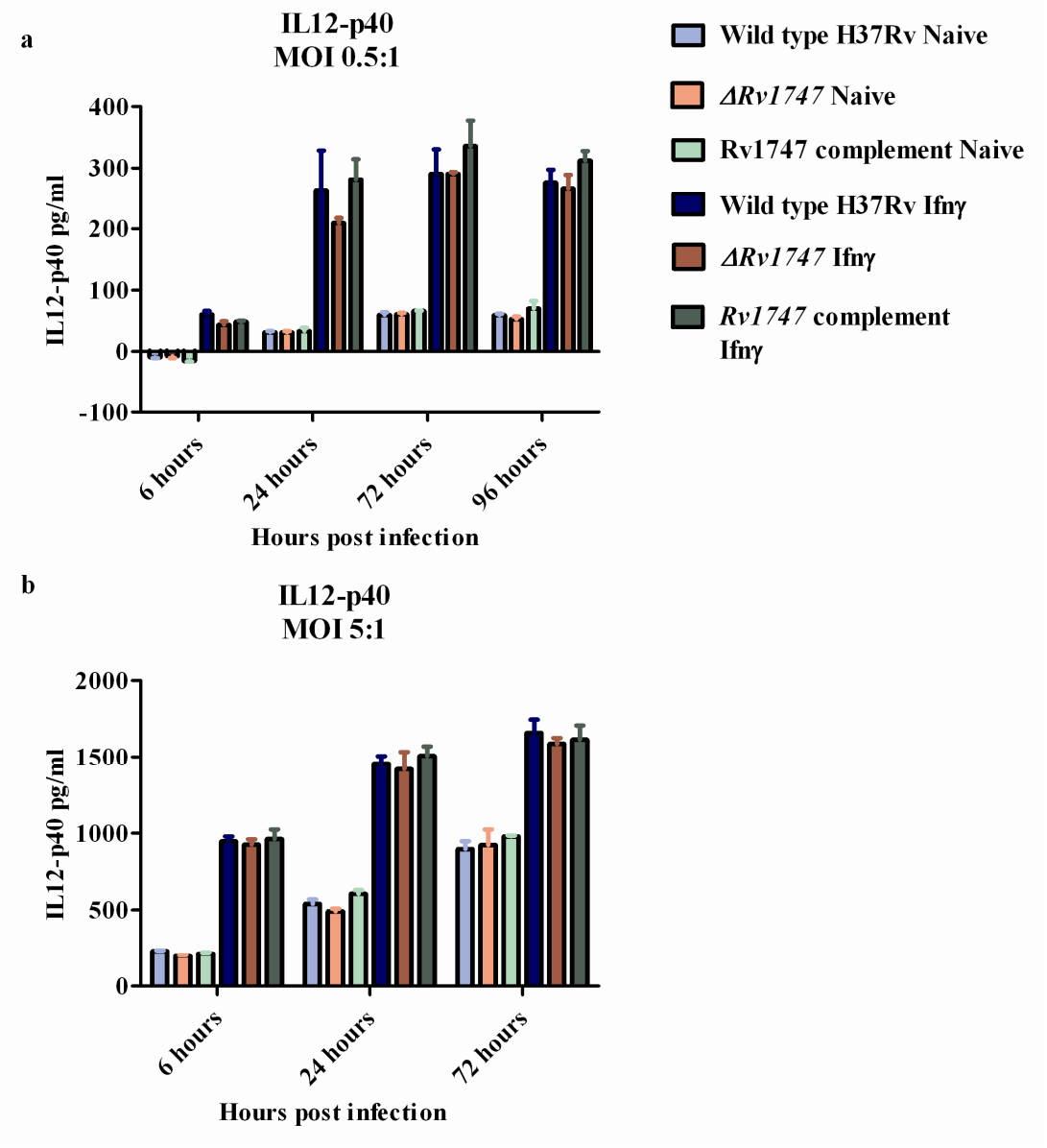


Figure 6.6 IL12-p40 production by BMDMs infected with wild type, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. Naïve and IFNγ activated BMDMs were infected with wild type H37Rv, *ΔRv1747* and *Rv1747* complement strains. At 6, 24, 72 (and 120 hours in figure 6.6a) hours post infection culture supernatants were prepared for ELISAs as described in section 2.4.14. Figure 6.6a shows the data generated from an *M. tuberculosis* infection with an MOI of 0.5:1 (*M. tb*:macrophage) with $2x10^5$ macrophages seeded per well. Figure 6.6b shows the data generated from an *M. tuberculosis* infection with an MOI of 5:1 (*M. tb*:macrophage) with $5x10^5$ macrophages seeded per well. Data plotted are the mean $\pm$ S.D of three technical replicates. Data show there is significantly more IL12-p40 produced by the IFNγ activated BMDMs compared with the naïve cells.

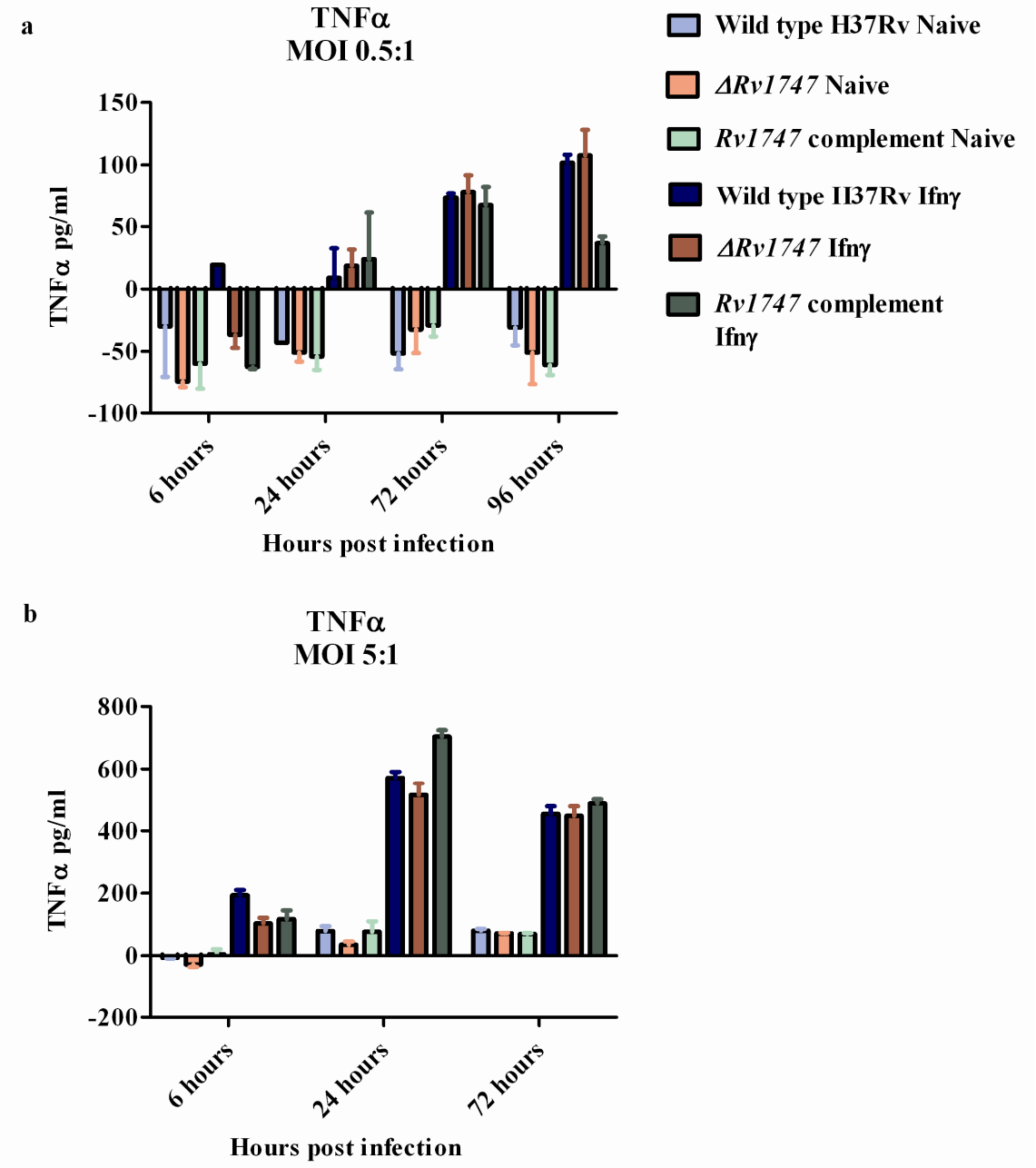


Figure 6.7 TNFα production by BMDMs infected with wild type, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. Naïve and IFNγ activated BMDMs were infected with wild type H37Rv, *ΔRv1747* and *Rv1747* complement strains. Figure 6.7a shows the data generated from an *M. tuberculosis* infection with an MOI of 0.5:1 (*M. tb*:macrophage) with $2x10^5$ macrophages seeded per well. Figure 6.6b shows the data generated from an *M. tuberculosis* infection with an MOI of 5:1 (*M. tb*:macrophage) with $5x10^5$ macrophages seeded per well. Data plotted are the mean $\pm$ S.D of three technical replicates. In all results there is significantly more TNFα produced by the IFNγ activated BMDMs compared with the naïve cells.

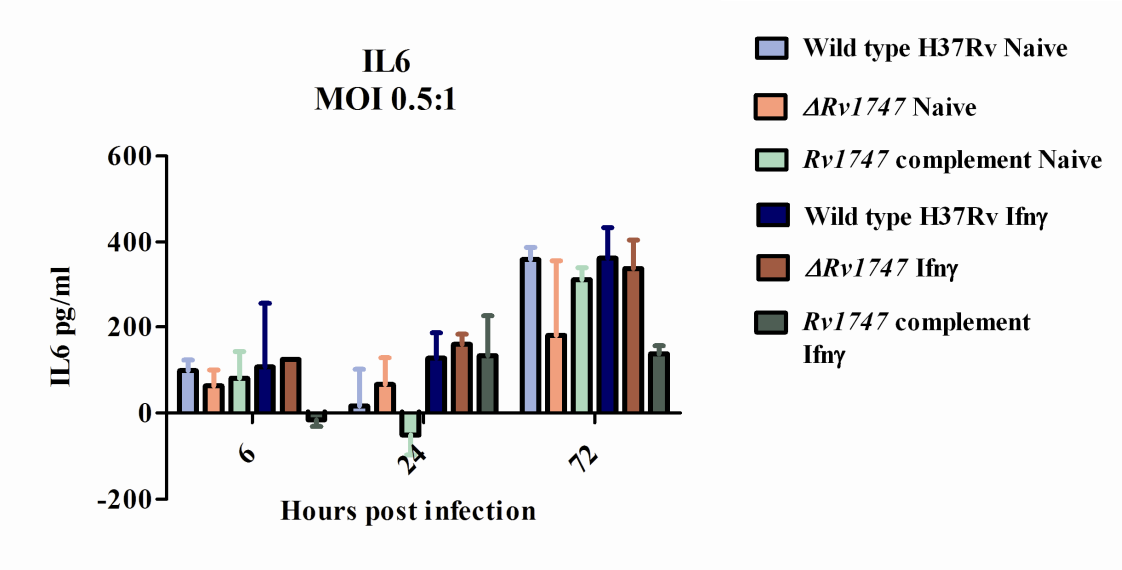


Figure 6.8 IL6 production by BMDMs infected with wild type, *ΔRv1747* and *Rv1747* complement *M. tuberculosis* strains. Naïve and IFNγ activated BMDMs were infected with wild type H37Rv, *ΔRv1747* and *Rv1747* complement strains. At 6, 24 and 72 hours post infection culture supernatants were prepared for ELISAs as described in section 2.4.14. Figure 6.8 shows the data generated from an *M. tuberculosis* infection with an MOI of 0.5:1 (*M. tb*:macrophage) with $2x10^5$ macrophages seeded per well. Data plotted are the mean $\pm$ S.D of three technical replicates.

complement strains. However, interestingly when the MOI was 0.5:1 at the 6 hour time point in IFNγ activated macrophages infected with the wild type strain the cells secreted 19 pg/ml of TNFα whereas macrophages infected with the mutant and complement strains produced levels below that of the negative control well (figure 6.7a). Furthermore, at 96 hours post infection the IFNγ activated cells secreted more TNFα when infected with the wild type and *ΔRv1747* strains (101 pg/ml and 107 pg/ml respectively) compared to the *Rv1747* complementing strain which secreted an average of 37 pg/ml (figure 6.7a).

Figure 6.8 shows the results of IL6 secretion into the culture supernatants. Results show that at all time points in both naïve and IFNγ activated macrophages there were no significant differences in the levels of IL6 secreted between the wild type and *ΔRv1747* strains. However, at 6 and 72 hours post infection in the IFNγ activated cells lower levels of IL6 production were calculated from the macrophages infected with the *Rv1747* complement strain compared to the wild type and deletion strains.

6.5 Assessment of the growth of the *ΔpknF* and *pknF* complement strains in a macrophage model of infection model

It was not known whether the *pknF* gene played a role in the growth of *M. tuberculosis in vivo*. This study to date had not discovered any observable phenotype in the *ΔpknF* null strain. Therefore the growth of the wild type, *ΔpknF* and *pknF* complement strains were assessed in the macrophage model of infection. Furthermore it was hypothesised that by examining the growth of this mutant in BMDMs it may also provide a further indication as to whether PknF positively or negatively regulates the activity of the Rv1747 ABC transporter.

The results from a BMDM infection experiment in both naïve (figure 6.9a) and IFNγ activated cells (figure 6.9b) showed that in naïve BMDMS there was no growth phenotype of the *ΔpknF* null or complement strains up until 168 hours post infection when a significant difference was observed in the bacterial load of the mutant and complement strains compared with the wild type strain; the *ΔpknF* and complement strains had a significantly lower bacterial load at this time point compared with the wild

type strain. Similarly, in IFNγ activated BMDMs the growth of all three strains was comparable up until 168 hours post infection where the *ΔpknF* strain had a significantly lower bacterial load compared with the wild type (the *pknF* complementing strain also had a lower bacterial load compared with the wild type but this was not found to be significant) (figure 6.9b).

Due to the differences in uptake of *M. tuberculosis* at six hours post infection the results were then normalised to the initial six hour uptake values of each strain (figure 6.10). Normalised values showed that the *pknF* complement had a similar growth pattern to the wild type strain; furthermore, the *ΔpknF* strain also followed a similar growth pattern to 120 hours post infection and then became attenuated for growth between 120 and 168 hours in naïve cells only ($p<0.05$).

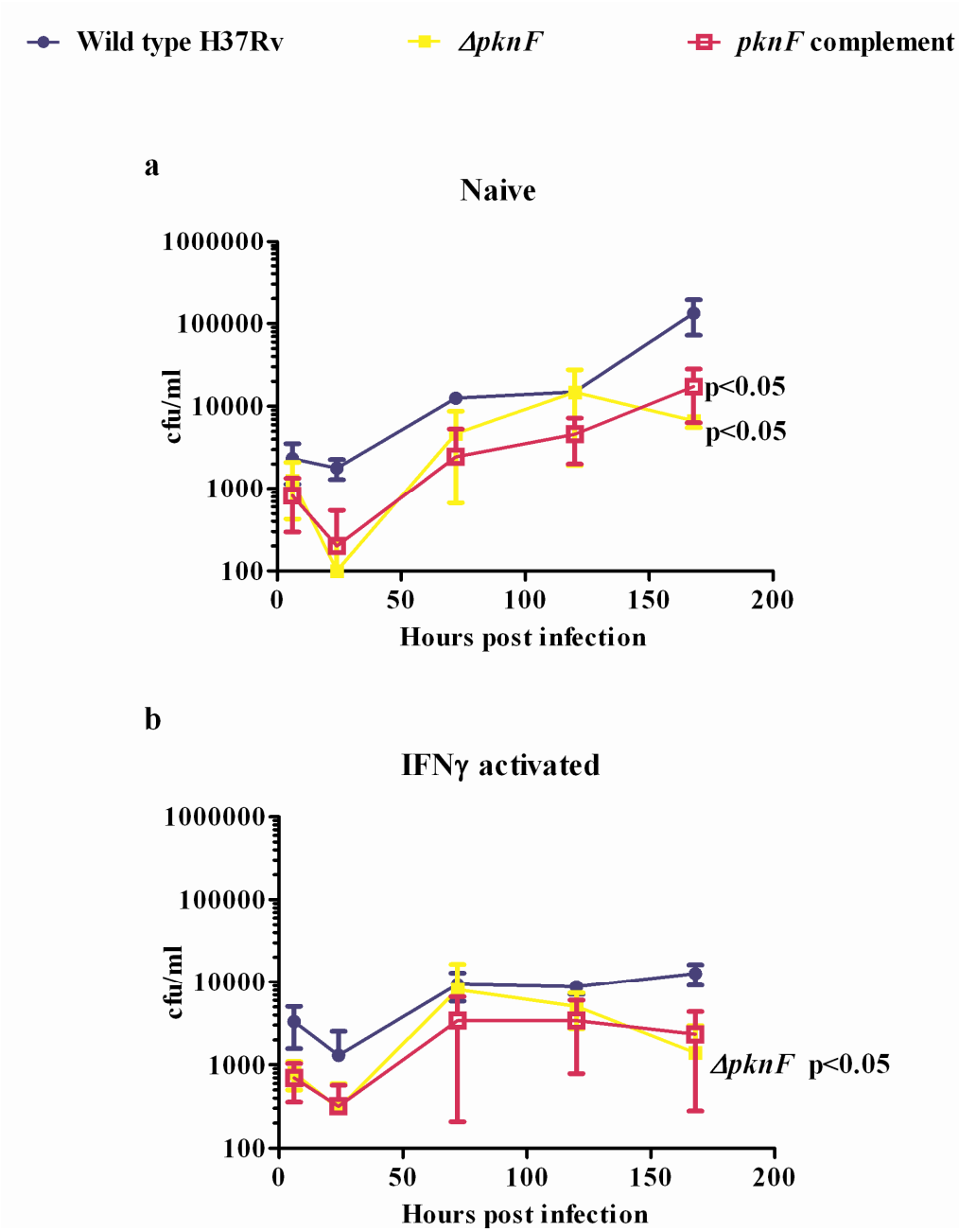


Figure 6.9 Bone marrow derived macrophage infection with wild type H37Rv, *ΔpknF* and *pknF* complement *M. tuberculosis* strains. Results show the growth of wild type (blue), *ΔpknF* (yellow) and *pknF* complement (pink) strains over 168 hours. Figure a shows the growth of *M. tuberculosis* strains in naïve macrophages and figure b shows the growth of *M. tuberculosis* strains in IFNγ activated cells. Data shown is a representative experiment from two biological replicates performed in triplicate. Error bars indicate the mean ±S.D of three technical replicates.

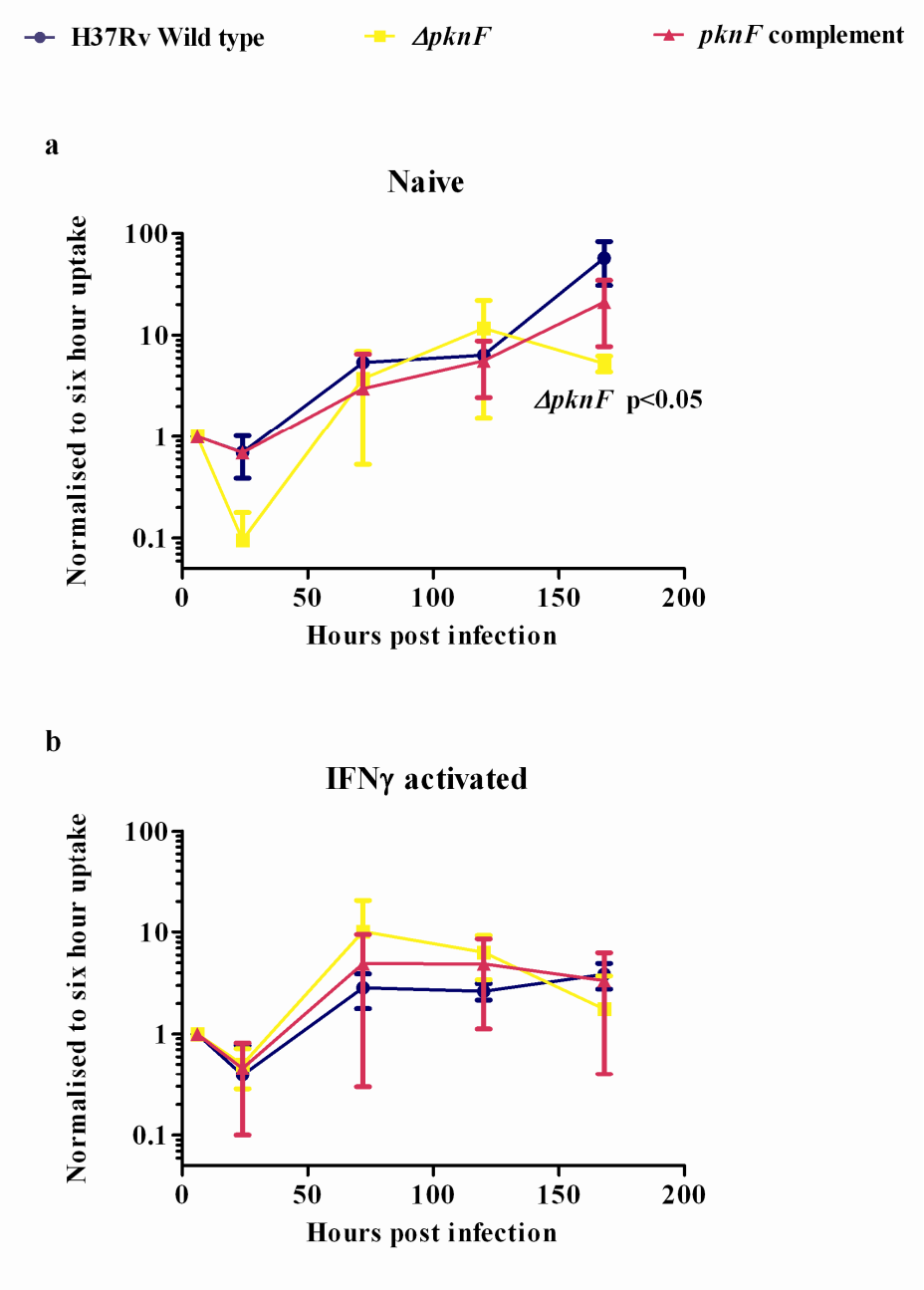


Figure 6.10 Bone marrow derived macrophage infection with wild type H37Rv, *ΔpknF* and *pknF* complement *M. tuberculosis strains* normalised to the six hour uptake data. Results show the normalised growth of wild type (blue), *ΔpknF* (yellow) and *pknF* complement (pink) strains over 168 hours. Figure a shows the growth of *M. tuberculosis* strains in naïve macrophages and figure b shows the growth of *M. tuberculosis* strains in IFNγ activated cells.

6.6 Characterisation of the function of the Rv1747 FHA domains

6.6.1 Assessment of the growth of the *Rv1747* complement S47A (ΔFHA-1) and *Rv1747* complement S248A (ΔFHA-2) strains in a murine macrophage infection model

The effect of rendering each FHA domain non-functional on bacterial growth was assessed using the macrophage infection model. It was hypothesised that if the function of either of the phospho-threonine binding domains were critical for Rv1747 function then a growth phenotype may be observed in this strain in a macrophage infection model; this may then have provided a clue as to the importance of each domain and may have helped to create a model of how the *pknF-Rv1747* signal transduction system is controlled. Site directed mutagenesis was used to create a serine to alanine mutation at position 47 of the FHA-1 domain and position 248 of the FHA-2 domain in the *Rv1747* complementing plasmid (R. Whalan, NIMR). The serine residue is essential for FHA domain binding to the pT epitopes so if either of the domains played an essential role in transduction of the signal to regulate the activity of the transporter then it was hypothesised that a phenotype may be observed in this mutant strain.

The results from a BMDM infection experiment in both naïve (figure 6.11a) and IFNγ activated cells (figure 6.11b) showed that the *Rv1747* complement S47A (FHA-1) strain was attenuated for growth compared with the wild type strain at 120 (IFNγ activated cells only) and 168 (naïve and IFNγ activated cells) hours post infection. However, the *Rv1747* complement S248A (FHA-2) strain showed a similar growth pattern to the wild type strain in both naïve and IFNγ activated macrophages. These results suggested that the FHA-1 domain is important for Rv1747 *in vivo* function.

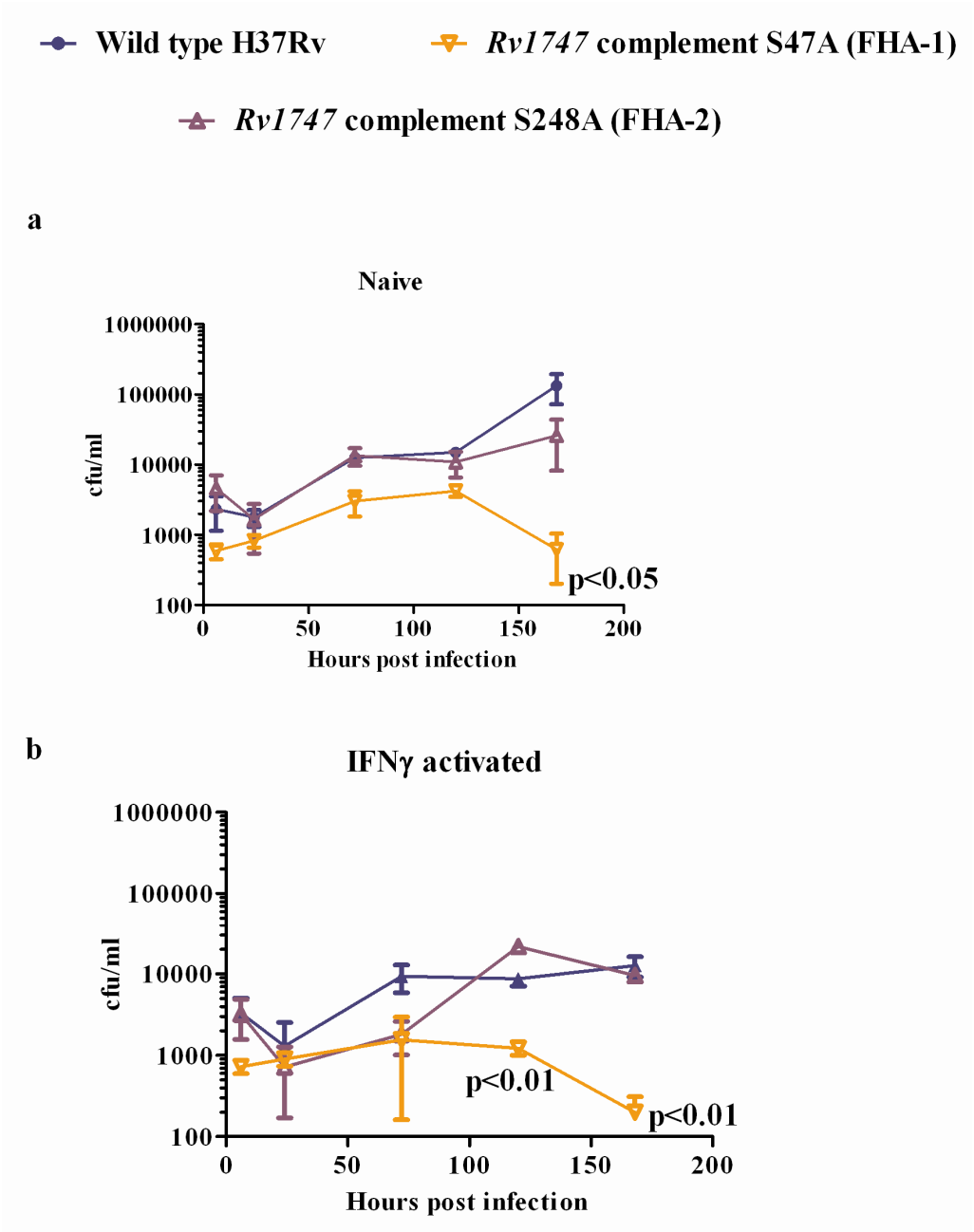


Figure 6.11 Bone marrow derived macrophage infection with wild type H37Rv, *ΔRv1747*, *Rv1747* complement S47A (ΔFHA-1) and *Rv1747* complement S248A (ΔFHA-2) *M. tuberculosis* strains. Infection of BMDMs with *M. tuberculosis* strains was performed with an MOI of 0.5:1 (*M.tb*:macrophage) as outlined in section 2.4.12.2. Results show the growth of wild type (blue), *Rv1747* complement S47A (ΔFHA-1) (orange) and *Rv1747* complement S248A (ΔFHA-2) (purple) strains over 168 hours. Figure a shows the growth of *M. tuberculosis* strains in naïve macrophages and figure b shows the growth of *M. tuberculosis* strains in IFNγ activated cells. Data show a representative experiment from two biological replicates performed in triplicate. Error bars indicate the mean ±S.D of three technical replicates. Data show that the *Rv1747* complement S47A (ΔFHA-1) strain is attenuated for growth in both naïve and activated macrophages.

6.6.2 Expression and purification of FHA-1, FHA-2 and FHA-1 S47A domains of Rv1747 from *E. coli*

The next aim of this study was to further elucidate the mechanism of how the FHA domains interacted with the pT peptides to regulate the activity of the Rv1747 ABC transporter. It was hypothesised that one of the FHA domains may be responsible for binding to one of the phospho-threonine epitopes and vice versa to regulate the control of the ABC transporter. If this hypothesis was correct then we may observe differences in the binding affinities between each FHA domain and the pT epitopes. The overall aim of this set of experiments was to further understand the function of the FHA domains by studying their interactions with the identified pT epitopes on Rv1747.

As a necessary preliminary in order to determine the binding kinetics between the FHA domains and the Rv1747 phosphopeptides each FHA domain first had to be expressed and purified from *E. coli* as described in section 2.5.2. Three protein constructs were purified; FHA-1 domain residues 1-120, FHA-2 domain residues 202-310 and FHA-1 S47A domain residues 1-120 (see figure 5.6 for a diagram highlighting the cloned residues). Site directed mutagenesis was performed to generate the S47A mutation within the FHA-1 domain construct. As previously mentioned FHA domains contain a conserved serine residue that is critical for binding to phosphopeptides hence this strain was generated to act as a control in the binding studies; an S47A mutation should render the FHA-1 domain incapable of binding the phosphopeptides. An SDS-PAGE gel shows each purified protein (figure 6.12).

The two identified phosphorylation sites on Rv1747 were synthesised as phosphopeptides by W. Mawby (Bristol University) for use in the binding studies with the FHA domains. Each peptide contained an identical N-terminal sequence which contained a tyrosine residue to allow for an accurate concentration to be determined by spectrometry. The sequence of the pT150 peptide was K-K-Y-A-G-Q-Q-L-P-P-A-pT-T-R-I-P-A-A and the sequence of the pT208 peptide was K-K-Y-A-G-T-E-A-G-N-L-A-pT-S-M-M-K where pT is the phosphothreonine residue.

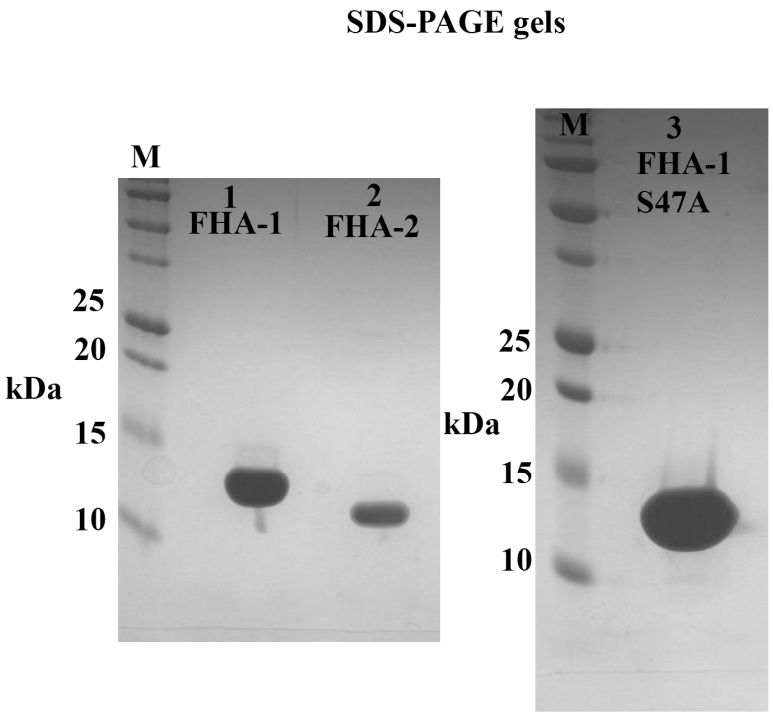


Figure 6.12 Expression and purification of FHA-1, FHA-2 and FHA-1 S47A domains of Rv1747 from *E. coli*. The three proteins were expressed and purified as described in section 2.5.2. After purification 10 µl of each pooled and concentrated protein sample were analysed by SDS-PAGE. Figure 6.12 shows the results of FHA-1 (lane 1), FHA-2 (lane 2) and FHA-1 S47A purification (lane 3) All proteins were purified to homogeneity. FHA-$1^{1\text{-}120}$ = 14 kDa, FHA-$2^{202\text{-}310}$ = 12 kDa.

6.6.3 Determination of the binding kinetics between the FHA-1 and FHA-2 domains with the Rv1747 pT150 and pT208 peptides

To determine the binding kinetics between both the FHA domains and the pT150 and pT208 phosphopeptides, isothermal titration calorimetry (ITC) was performed as described in section 2.5.3.3. The aim of ITC was to calculate the dissociation constant (K_d) which measures how tightly a ligand binds to a particular protein. The K_d is defined as the concentration of ligand at which the binding site on a particular protein is half occupied, hence the smaller the dissociation constant, the more tightly bound the ligand. In the following experiments the aim was to calculate the K_d between each FHA domain and the pT peptides.

An ITC trace consists of two panels; the upper panel plots the raw data for an experiment which consisted of a series of heat spikes plotted as the amount of power (in μcal/sec) needed to maintain the reference and sample cell at the same temperature with each spike representing one injection of ligand. The lower panel showed the total heat effect per injection by integrating the area of the peaks and plotting then against the molar ratio of protein and ligand. A curve was then fitted to the data and this generated the binding data. From the shape of the curve it was possible to determine the stochiometry and enthalpy of the reaction as well as the binding affinity, the inverse of which is the dissociation constant.

The results of the analysis of binding of the FHA-1 domain to the Rv1747 pT150 (figure 6.13a) and pT208 phosphopeptides (figure 6.13b) demonstrated that the K_d was 20 μM between FHA-1 and pT150 and 16 μM between FHA-1 and pT208. The results of the analysis of binding of the FHA-2 domain to the Rv1747 pT150 (figure 6.14a) and pT208 phosphopeptides (figure 6.14b) calculated that the K_d was 16 μM between FHA-2 and pT150 and 18 μM between FHA-2 and pT208.

The results of the analysis of binding of the FHA-1 S47A domain to the Rv1747 pT150 (figure 6.15a) and pT208 phosphopeptides (figure 6.15b) confirmed that mutation of the critical serine residue within the FHA-1 domain responsible for binding to the pT

residue completely abolished binding to both the pT peptides. This result showed that the wild type FHA domains were indeed specifically binding to the pT peptides.

6.6.4 Determination of the binding kinetics between the FHA-1 and FHA-2 domains with the Rv1827 pT22 peptides

Both the Rv1747 FHA domains were assessed for their ability to bind an Rv1827 pT22 phosphopeptide. This investigation was performed to establish whether there were any notable differences in the ability of the Rv1747 FHA domains to bind a non-Rv1747 phosphopeptide compared with the Rv1747 peptides. The sequence of the synthesised Rv1827 pT22 peptide was E-V-T-V-E-T-pT-S-V-F-R-A-D-Y-K-K where pT is the phosphothreonine residue. The results of the analysis of binding of the FHA-1 (figure 6.16a) and FHA-2 (figure 6.16b) domains to the Rv1827 pT22 phosphopeptide calculated that the K_d was 22 μM between FHA-1 and pT22 and 21 μM between FHA-2 and pT22. The calculated dissociation constants were quite similar to those found in the experiments which determined the binding kinetics between the FHA domains and the Rv1747 phosphopeptides.

6.6.5 Determination of the binding kinetics between the FHA-2 domain and the Rv0020c pT377 peptide

It was then assessed whether the FHA-2 domain could bind a phosphothreonine epitope from Rv0020c which along with Rv1827 is also phosphorylated by PknB (O'Hare *et al.*, 2008; Villarino *et al.*, 2005). The sequence of the synthesised Rv0020c pT377 peptide was A-D-T-pT-R-Y-T-E-S. The results of the binding of the FHA-2 domain to the Rv0020c pT277 peptide (figure 6.17) showed that the K_d was 9 μM between FHA-2 and Rv0020c pT377. Again, the calculated dissociation constant was quite similar to those calculated in the experiments which determined the binding kinetics between the FHA domains and the Rv1747 phosphopeptides.

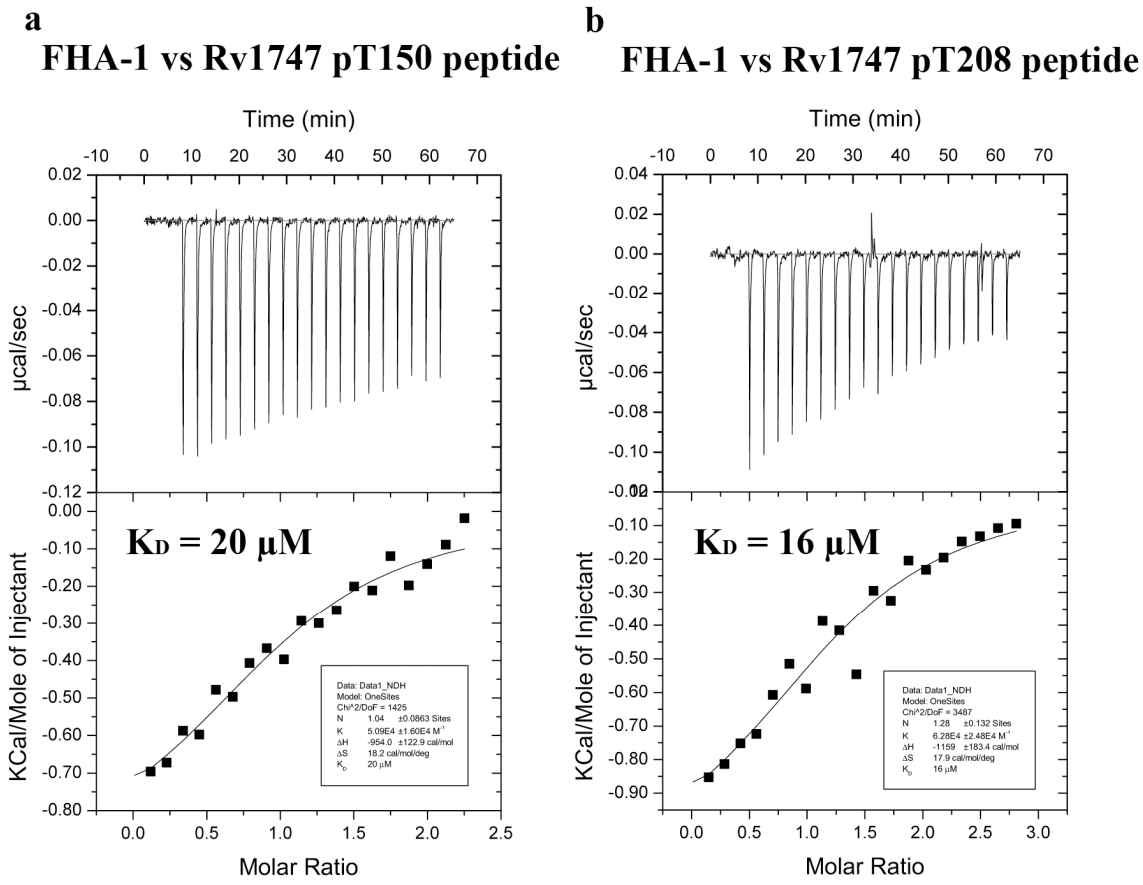


Figure 6.13 Analysis of binding of Rv1747 FHA-1 domain to the Rv1747 pT150 and pT208 phosphopeptides. ITC was used to determine the binding kinetics between FHA domains and potential phosphopeptide substrates and was performed as described in section 2.5.3.3. Figure 6.13a shows the binding of the FHA-1 domain to the pT150 phosphopeptide and figure b shows the binding of the FHA-1 domain to the pT208 phosphopeptide. The K_d was calculated as 20 µM between FHA-1 and pT150 and 16 µM between FHA-1 and pT208.

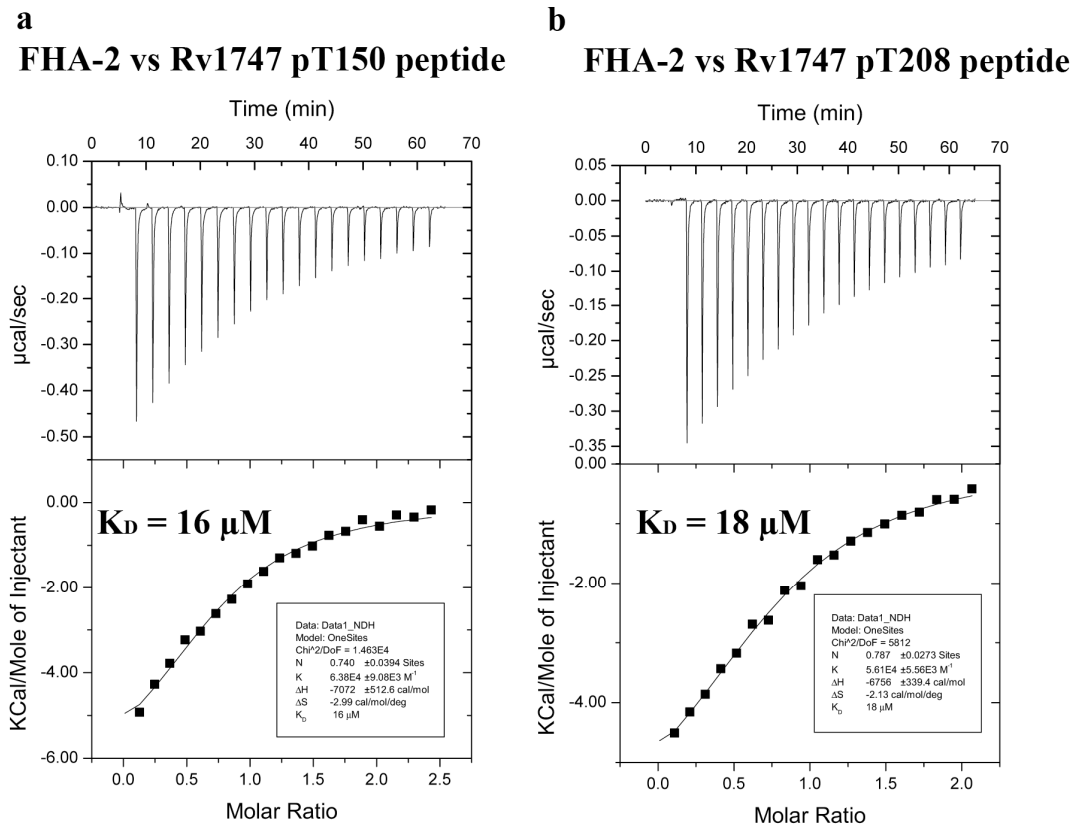


Figure 6.14 Analysis of binding of Rv1747 FHA-2 domain to the Rv1747 pT150 and pT208 phosphopeptides. Figure 6.14a shows the binding of the FHA-2 domain to the pT150 phosphopeptide and figure b shows the binding of the FHA-2 domain to the pT208 phosphopeptide. The K_d was calculated as 16 µM between FHA-2 and pT150 and 18 µM between FHA-2 and pT208.

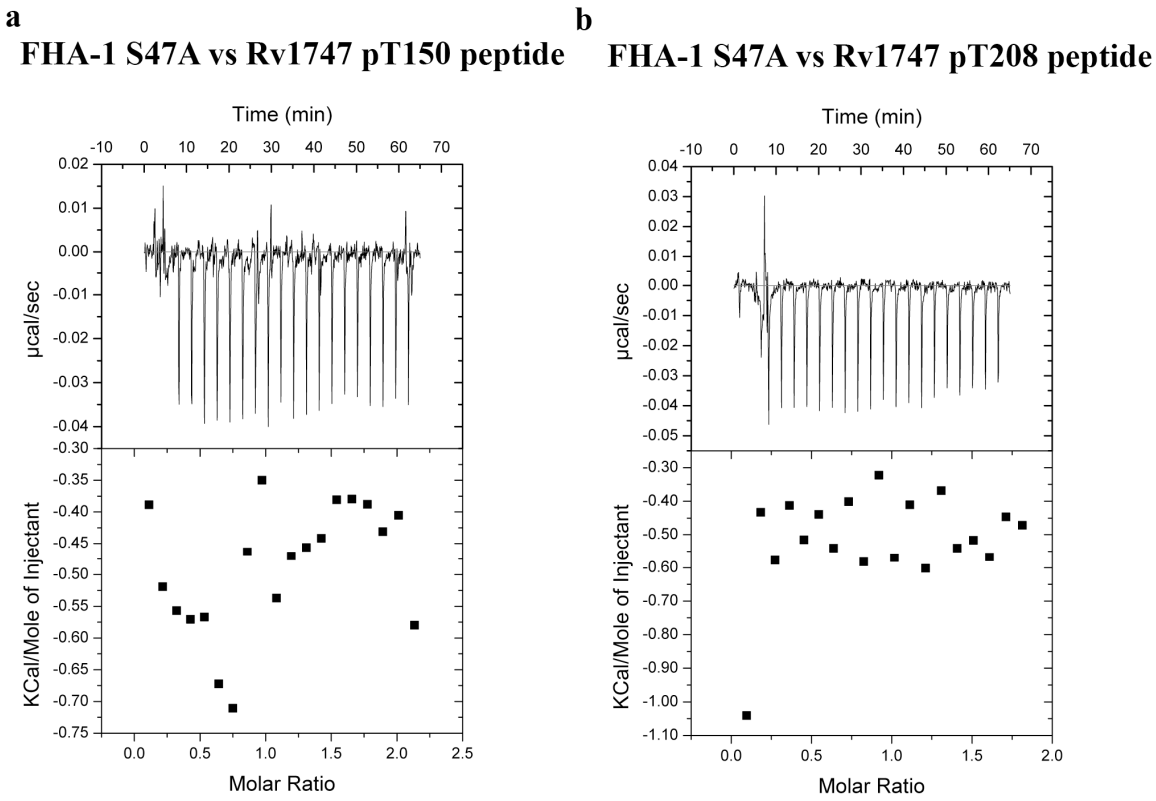


Figure 6.15 Analysis of the binding of the Rv1747 FHA-1 S47A domain to the Rv1747 pT150 and pT208 phosphopeptides. Figure 6.15a shows the binding of the FHA-1 S47A domain to the pT150 phosphopeptide and figure b shows the binding of the FHA-1 S47A domain to the pT208 phosphopeptide. Results show that FHA-1 S47A cannot bind the pT150 or pT208 peptides. Note the very small scale in the top panels of figures a and b; the heat spikes generated were due to the heat of dilution that occurs when two protein species are mixed, this occurrence is not associated with protein binding.

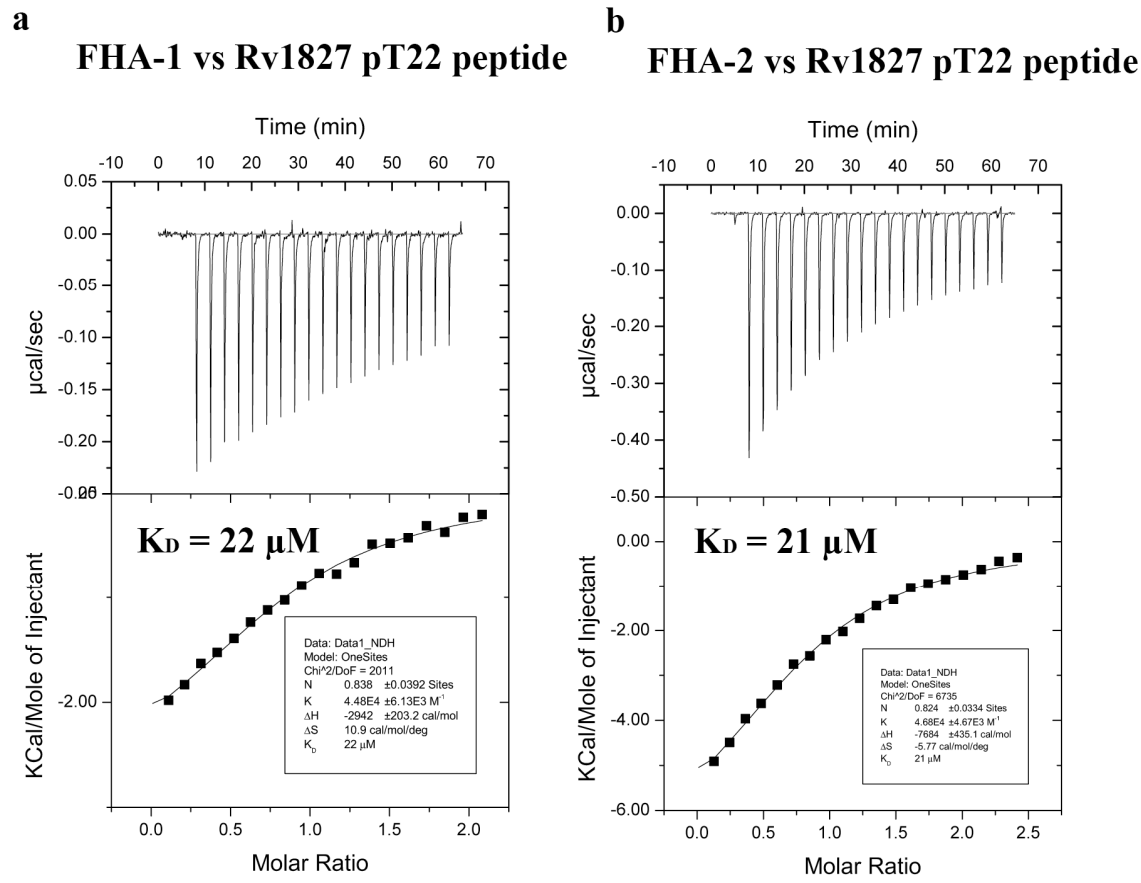


Figure 6.16 Analysis of binding of the Rv1747 FHA-1 and FHA-2 domains to the Rv1827 pT22 phosphopeptide. Figure 6.16a shows the binding of the FHA-1 domain to the Rv1827 pT22 phosphopeptide and figure b shows the binding of the FHA-2 domain to the Rv1827 pT22 phosphopeptide. The K_d was calculated as 22 µM between FHA-1 and Rv1827 pT22 and 21 µM between FHA-2 and Rv1827 pT22.

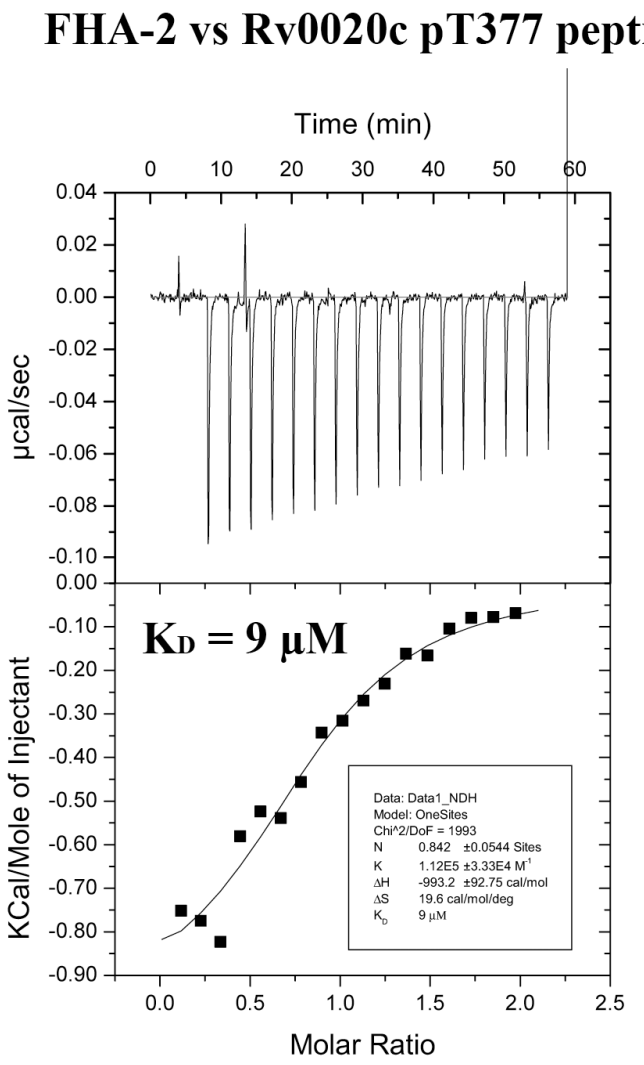


Figure 6.17 Analysis of binding of the Rv1747 FHA-2 domain to the Rv0020c pT377 phosphopeptide. The figure shows the binding of the FHA-2 domain to Rv0020c pT377 phosphopeptide. The K_d was calculated as 9 µM between FHA-2 and Rv0020c pT377.

6.7 Does the FHA-2 domain dimerise upon phosphorylation at T208?

Due to the close proximity of the pT208 residue to the FHA-2 domain (figure 5.6) it would not be possible for the FHA-2 domain to rotate in such a way that the phosphothreonine binding domain could bind the pT208 residue in its peptide binding pocket. Therefore, a hypothesis was proposed that upon T208 phosphorylation the FHA-2 domain on one half of the transporter may bind the T208 residue on the other half and vice versa in order to transduce the signal and regulate the activity of the Rv1747 transporter as depicted in figure 6.18; this hypothesis is similar to what was shown for the Chk2 protein (Li *et al.*, 2008).

The next set of experiments performed in this study set out to determine whether FHA-$2^{202-310}$ became dimeric after phosphorylation at position T208 by $PknF^{1-292}$.

6.7.1 Determination of the phosphorylation of FHA-2 domain by PknF

Prior to studies being carried out to determine whether FHA-$2^{202-310}$ becomes dimeric upon $PknF^{1-292}$ phosphorylation it first had to be assessed whether $PknF^{1-292}$ could phosphorylate FHA-$2^{202-310}$ *in vitro*.

6.7.1.1 *In vitro* phosphorylation of Rv1747, FHA-1 and FHA-2 domain by PknF

An *in vitro* phosphorylation assay showed that $PknF^{1-292}$ was capable of phosphorylating the FHA-1^{1-120} domain (figure 6.19b, lane 3) and most importantly there was a large radioactive band corresponding to the phosphorylation of FHA-$2^{202-310}$ domain by PknF (figure 6.19b, lane 4). These results showed that PknF could indeed phosphorylate the FHA-2 domain *in vitro*. PknF could also phosphorylate the FHA-1 domain construct but to a much lesser extent (figure 6.19b, lane 3). This is because within the FHA-2 construct the pT208 site is present whereas within the FHA-1 construct no identified pT site was present (see figure 5.6 for diagram of cloned residues). However, the presence of a radioactive band corresponding to phosphorylated FHA-1 indicated that this protein domain could still be phosphorylated to some extent and there are several threonine residues within this construct on which phosphorylation could occur.

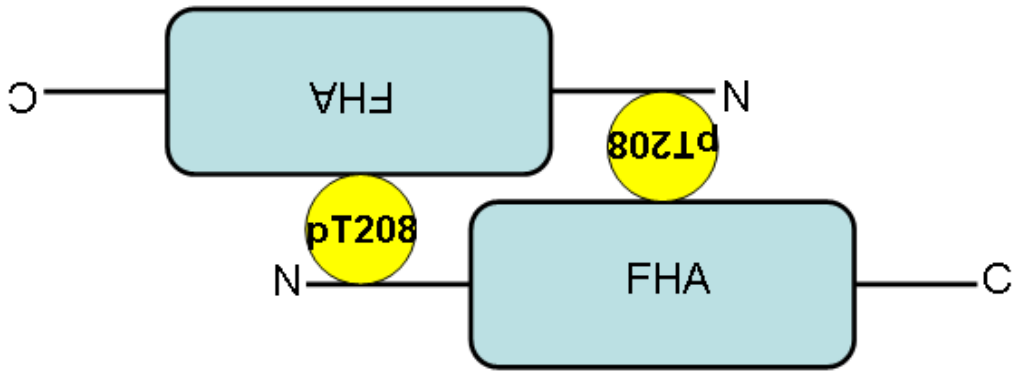


Figure 6.18 Schematic representation of how FHA-2$^{202-310}$ could potentially engage in an intermolecular interaction after phosphorylation at residue T208. The schematic shows how upon Rv1747 phosphorylation by PknF at residue T208 FHA-2 could potentially engage in an intermolecular interaction and become dimeric. N-terminal phosphothreonine residues are represented by a yellow circle containing pT208. N and C represent the N- and C- termini of the protein.

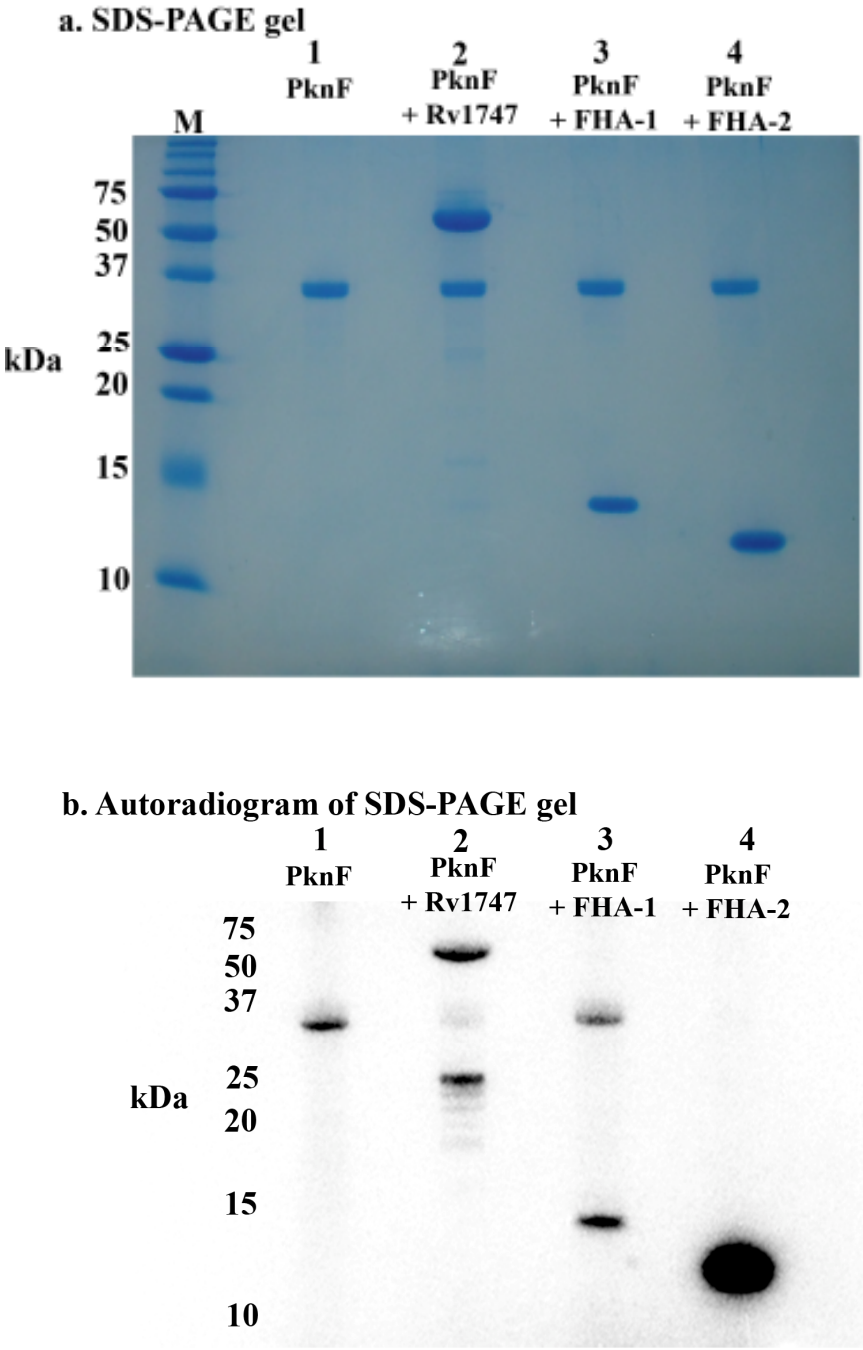


Figure 6.19 *In vitro* phosphorylation of Rv1747, FHA-1 and FHA-2 by PknF. *In vitro* phosphorylation assays were performed with the proteins as outlined in section 2.5.3.1. The proteins were incubated with 200 µCi/ml [γ-^{32}P] ATP then subjected to gel electrophoresis (Figure 6.19a) and radioactive bands were revealed by autoradiography (Figure 6.19b). Approximately 1 µg of protein was loaded into each lane. Figure 6.19b shows that PknF$^{1-292}$ is an active kinase capable of autophosphorylation (lane 1) which phosphorylates Rv1747$^{1-559}$ (lane 2), FHA-1$^{1-120}$ (lane 3) and in particular FHA-2$^{202-310}$ (lane 4). PknF$^{1-292}$ = 32 kDa, Rv1747$^{1-559}$ = 58 kDa, FHA-1$^{1-120}$ = 14 kDa, FHA-2$^{202-310}$= 12 kDa.

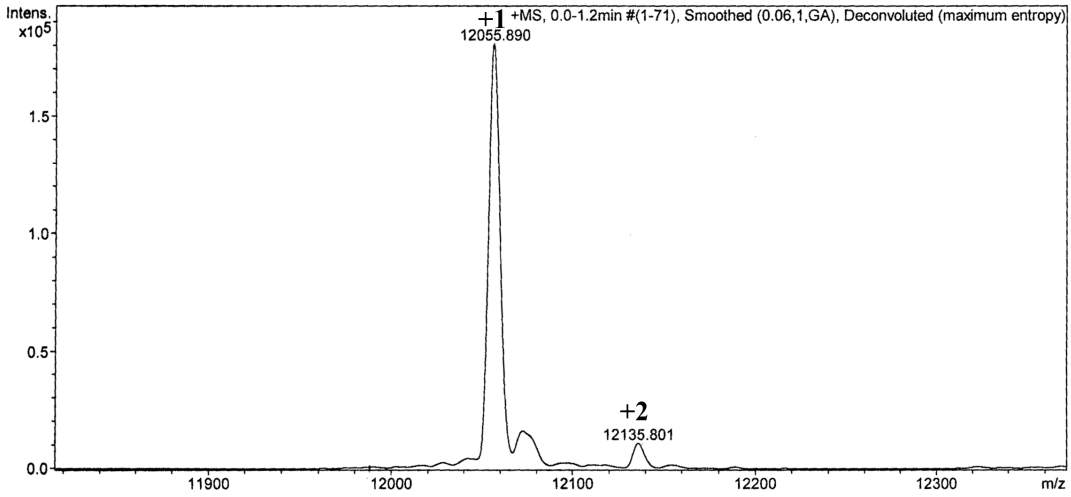
6.7.1.2 Use of electrospray mass spectrometry to determine the phosphorylation status of FHA2$^{202-310}$

A condition then had to be determined where the FHA-2 domain was phosphorylated by PknF in preparation for SEC-MALLS. The addition of one additional phosphate group was assessed using electrospray mass spectroscopy. PknF, FHA-2 and ATP were incubated together as described in section 2.5.3.4 before being assessed for the incorporation of one additional phosphate group on the FHA-2$^{202-310}$ domain. The results of electrospray mass spectroscopy after four hours of incubation showed that almost all the FHA-2 domain construct contained one additional phosphate group (figure 6.20). No protein remained which was not phosphorylated and there was a very small amount of the overall protein which had gained two additional phosphate groups. It had to be assumed that the additional phosphate group was added on to residue T208 of the FHA-2$^{202-310}$ domain construct.

Taken together, these results showed that PknF$^{1-292}$ was indeed capable of phosphorylating FHA-2$^{202-310}$ domain *in vitro* and by mass spectrometry it was demonstrated that a four hour incubation with ATP was sufficient to add one additional phosphate group to the protein construct. This protein, along with a non-phosphorylated FHA-2$^{202-310}$ control was then assessed by SEC-MALLS to investigate whether phosphorylated FHA-2 was indeed dimeric.

6.7.2 Use of SEC-MALLS to determine if the FHA-2 domain dimerises upon phosphorylation

SEC-MALLS is a technique which combines size exclusion chromatography and light scattering to measure the absolute molecular mass and concentration of proteins of interest. This technique was employed to determine the conformation of the FHA-2$^{202-310}$ domain after phosphorylation by PknF$^{1-292}$ and was performed as described in section 2.5.3.4.



FHA2 202-310 computed Mw = 11976.5

Figure 6.20 Electrospray mass spectrometry of FHA-2$^{202-310}$ phosphorylated for four hours with PknF$^{1-292}$. Each peak labelled '+' followed by a number represents the number of additional phosphate groups present on the FHA-2 construct. The net gain of a phosphate group results in an increase in mass of 80 Da. Data show that after four hours of phosphorylation with PknF non-phosphorylated FHA-2 could not be detected and all the FHA-2 protein contained one additional phosphate group. In addition, a very small proportion of the protein contained two additional phosphate groups.

Figure 6.21 shows the results of a SEC-MALLS experiment. The traces showing phosphorylated FHA-$2^{202-310}$ (red traces in figure 6.21) were compared with those showing the non-phosphorylated FHA-$2^{202-310}$ domain (blue traces in figure 6.21). The results show that PknF eluted between 28 and 31 minutes (figure 6.21b) with a confirmed molecular weight of approximately 32 kDa (the correct size for recombinant $PknF^{1-292}$). The results also show that phosphorylated FHA-2 eluted from the column prior to non-phosphorylated FHA-2; there was a difference in elution times of approximately 1.75 minutes. The fact that the phosphorylated FHA-2 protein eluted before the non-phosphorylated version suggested that the phosphorylated version might be larger.

However, it can be observed that PknF was still eluting from the column when the FHA-2 proteins (especially the phosphorylated version) began to elute (figure 6.21b). Consequently it was not possible to determine a precise molecular weight of the phosphorylated version of FHA-2.

Therefore an alternative strategy was sought; it was decided to express and purify GST-$PknF^{1-292}$ and phosphorylate the FHA-2 construct with this protein; the GST tag is 26 kDa in size meaning that the kinase and the FHA-2 domain should then elute separately. What is more, GST is dimeric meaning that the GST-PknF construct would actually run as a ~116 kDa protein.

On figure 6.21a the peak eluting between 10 and 20 minutes is when the void volume had run through the column and the first protein species eluted. This fraction may have included protein contaminants and any large complexes in the reactions; large complexes eluting after the void volume can also be seen in figures 6.23a and 6.24a.

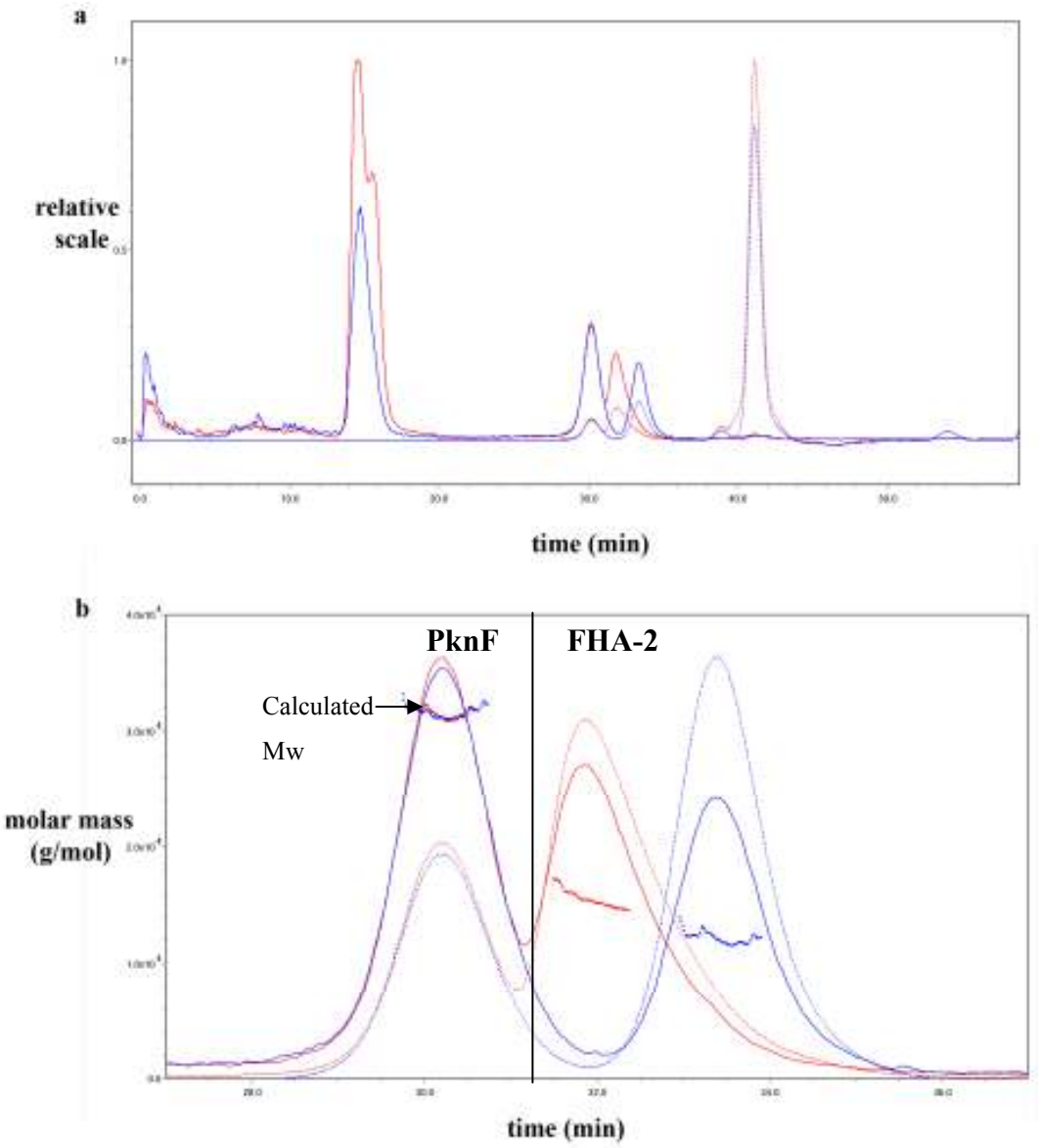


Figure 6.21 SEC-MALLS of FHA-2 domain with and without phosphorylation by PknF. SEC-MALLS was performed as described in section 2.5.3.4. Figure 6.21a shows the entire size exclusion experiment and figure 6.21b focuses on the proteins which eluted between 27 and 37 minutes. The protein which eluted between 28 and 31 minutes in both the phosphorylated and non-phosphorylated experiment was PknF with a confirmed molecular weight of approximately 32 kDa. Figure 6.21b shows that phosphorylated FHA-2 eluted from the column before non-phosphorylated FHA-2.

6.7.3 Expression and purification of GST-PknF$^{1-292}$ from *E. coli*

GST-PknF$^{1-292}$ was expressed and purified; the results of the protein purification demonstrated that the protein was purified to near homogeneity (figure 6.22). The GST-tagged kinase was then used to phosphorylate FHA-2$^{202-310}$ to determine whether this phospho-threonine binding domain became dimeric upon T208 phosphorylation.

6.7.4 SEC-MALLS of FHA-2 phosphorylated by GST-PknF$^{1-292}$

Figure 6.23 shows the results of SEC-MALLS of the FHA-2 domain without phosphorylation. The figure shows that FHA-2 at a concentration of 1 mg/ml and FHA-2 at a concentration of 4 mg/ml have identical elution profiles and the calculated molecular mass of the proteins are the same. The traces of the phosphorylated FHA-2 domains were then compared. Three different concentrations of phosphorylated FHA-2 were compared; 1 mg/ml, 2 mg/ml and 4 mg/ml (figure 6.24). The results showed that the calculated molecular weights of all three concentrations of protein were comparable confirming that increasing protein concentration does not alter the elution profile.

Figure 6.25 shows an overlay of the elution profiles of both the non-phosphorylated FHA-2 domain and the phosphorylated FHA-2 domain. The results show that the phosphorylated FHA-2 domain began to elute approximately 2 minutes prior to the non-phosphorylated FHA-2 domain. The molecular weight of the phosphorylated FHA-2 domain was shown to be 14.940 kDa at 31.8 minutes and the non-phosphorylated FHA-2 domain had a calculated mass of 10.940 kDa at 33.5 minutes.

The results showed that the phosphorylated FHA-2 domain elutes prior to the non-phosphorylated FHA-2 domain suggesting that the phosphorylated domain is larger; however, the calculated molecular weights of the two proteins does not provide an insight into whether the FHA-2$^{202-310}$ domain alters its conformation upon phosphorylation.

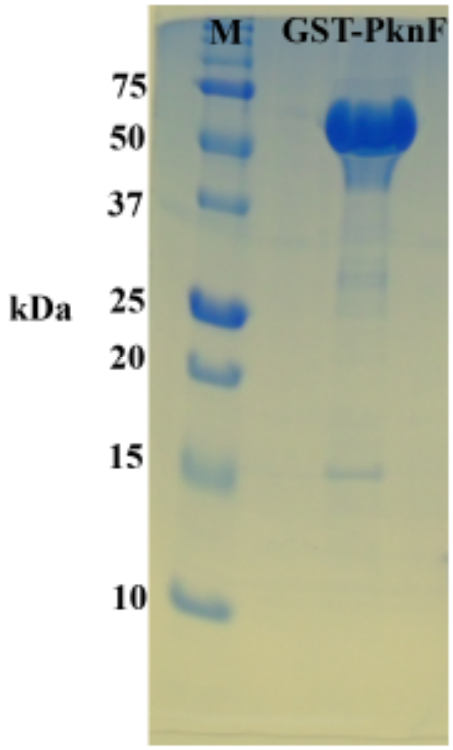


Figure 6.22 Expression and purification of GST-PknF$^{1-292}$ from *E. coli*. GST-PknF$^{1-292}$ was expressed and purified as described in section 2.5.2. After purification 10 µl of the pooled and concentrated protein sample was analysed by SDS-PAGE; figure 6.22 shows the result of a GST-PknF$^{1-292}$ purification. GST-PknF$^{1-292}$ = 58 kDa.

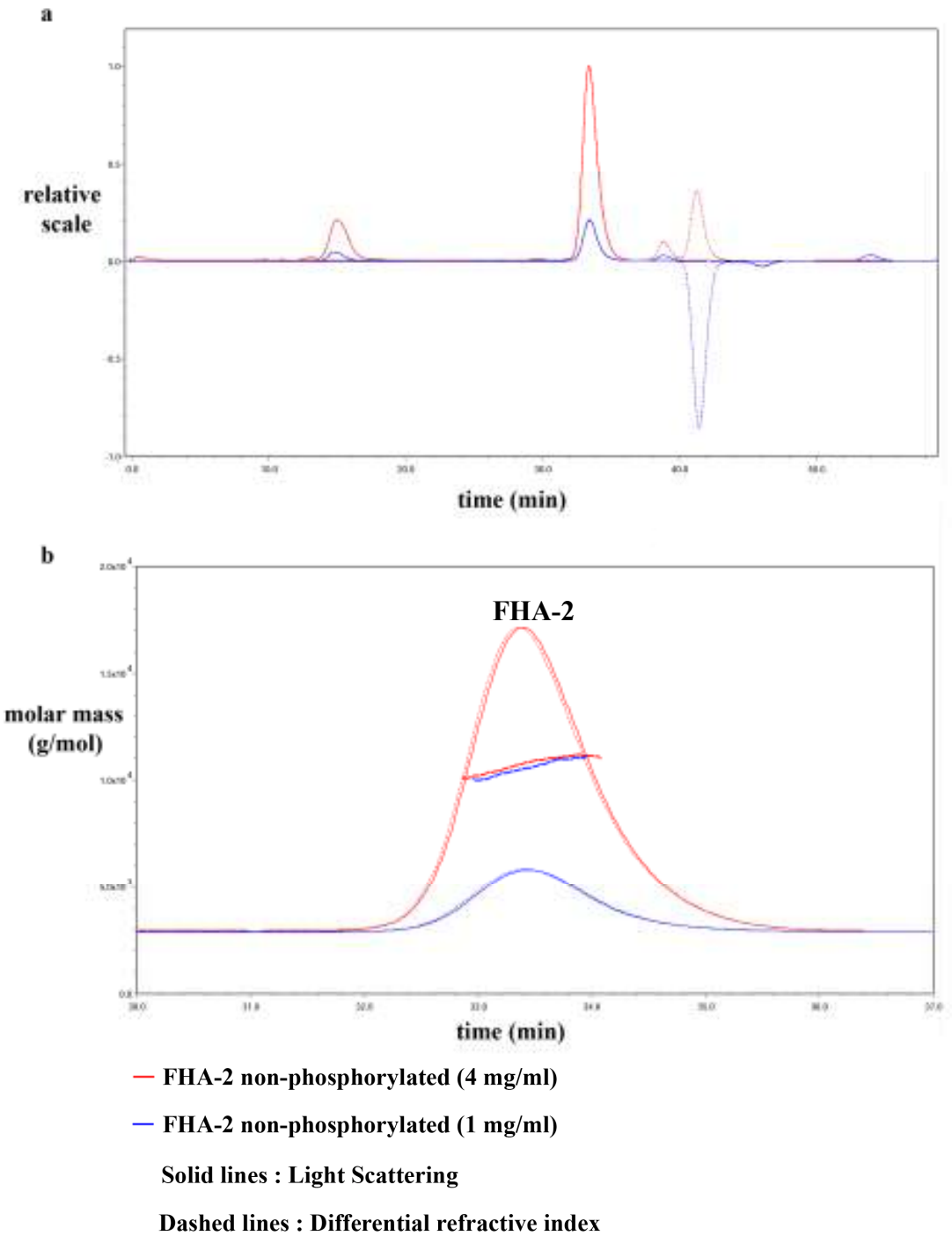


Figure 6.23 SEC-MALLS of the FHA-2 domain without phosphorylation by GST-PknF. Figure 6.23a shows the entire size exclusion experiment and figure 6.23b focuses on the proteins which eluted between 30 and 37 minutes. The figure shows that FHA-2 at a concentration of 1 mg/ml and FHA-2 at a concentration of 4 mg/ml have identical elution profiles and the calculated molecular mass of the proteins are the same.

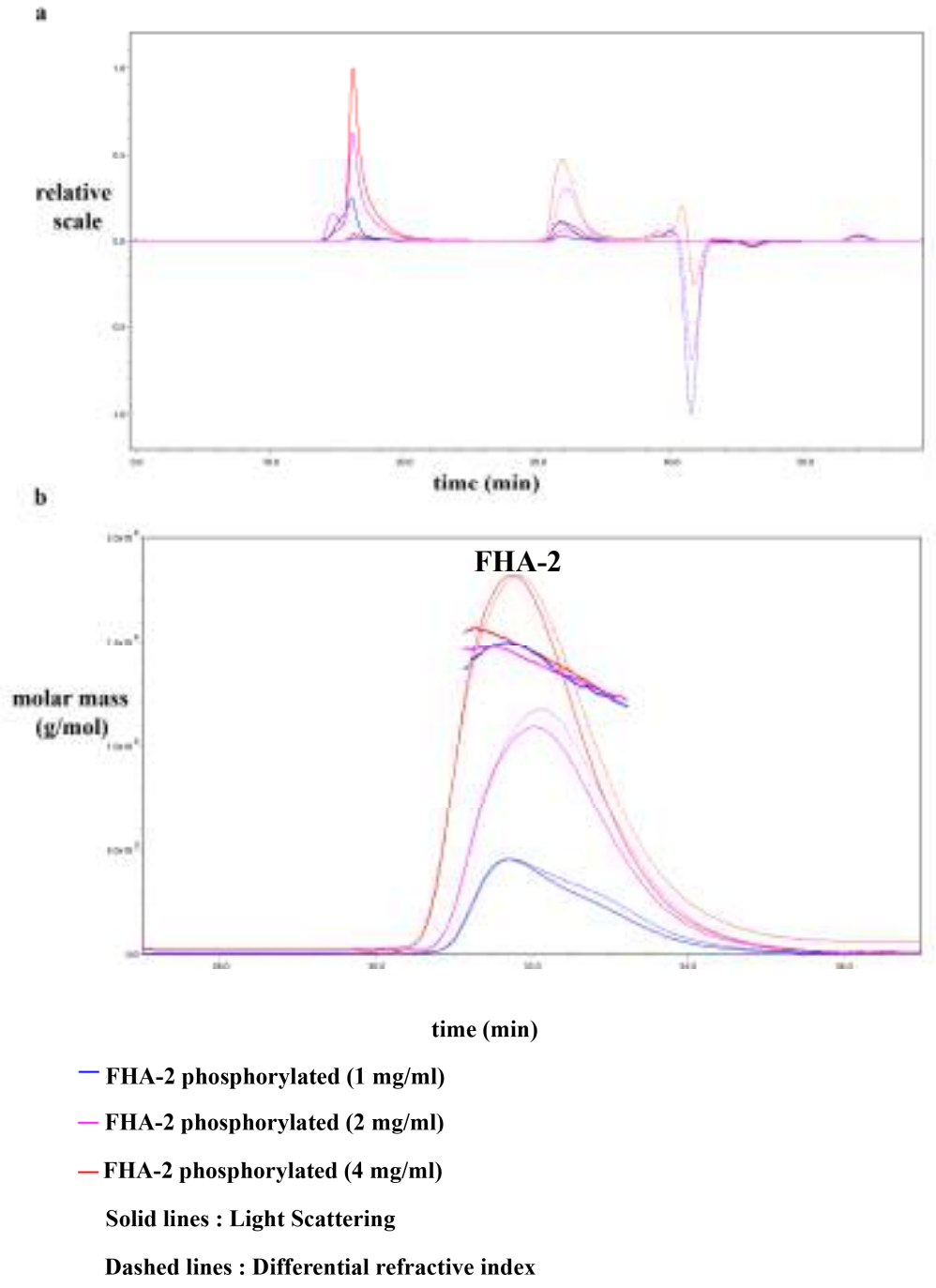


Figure 6.24 SEC-MALLS of the FHA-2 domain after phosphorylation by GST-PknF. Three concentrations of the FHA-2 construct were phosphorylated by GST-PknF$^{1-292}$ and then SEC-MALLS was performed. Figure 6.24a shows the entire size exclusion experiment and figure 6.24b focuses on the proteins which eluted between 27 and 37 minutes. FHA-2 at a concentration of 1 mg/ml, 2 mg/ml and 4 mg/ml are shown by the blue, pink and red traces respectively. The figure shows that the molecular mass of the three different concentrations of protein are approximately the same.

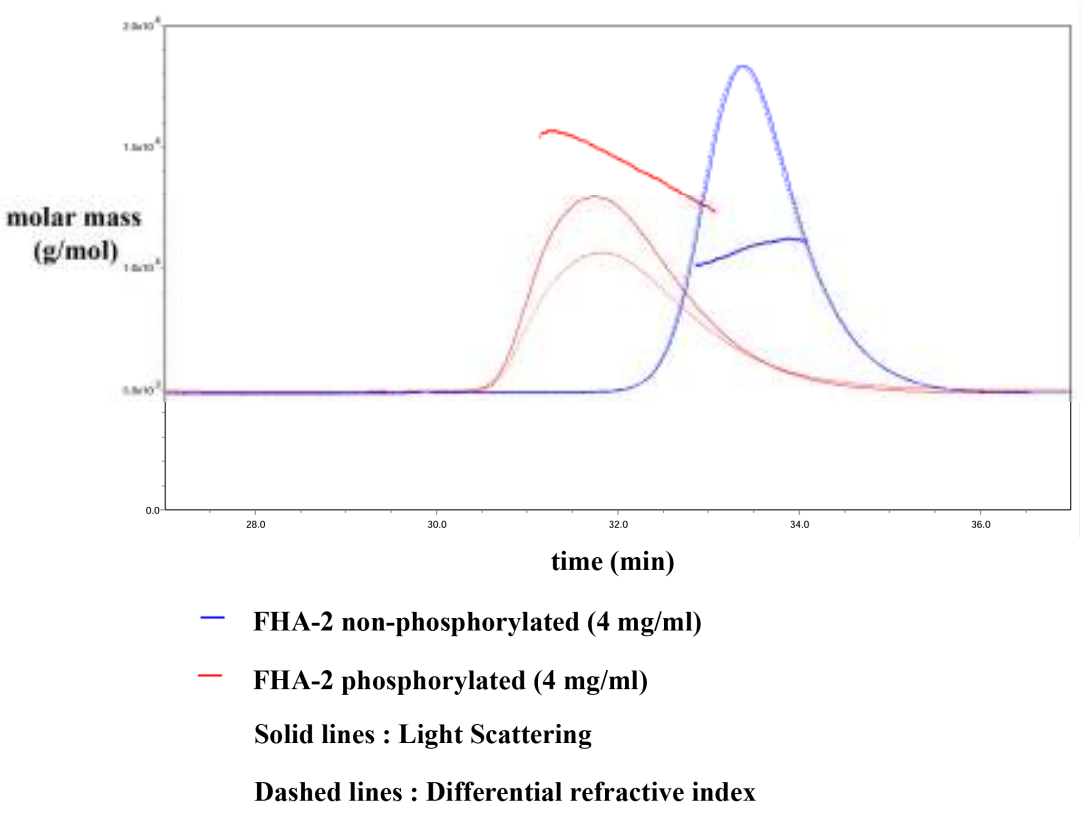


Figure 6.25 SEC-MALLS of the FHA-2 domain with and without phosphorylation by GST-PknF. The FHA-2 construct was phosphorylated by GST-PknF$^{1-292}$ when required and then SEC-MALLS was performed. Figure 6.25 shows an overlay of figures 6.23b and 6.24b with 4 mg/ml of FHA-2 domain protein. The results show that the phosphorylated FHA-2 domain begins to elute approximately 2 minutes prior to the non-phosphorylated FHA-2 domain. The molecular weight of the phosphorylated FHA-2 domain was shown to be 14.940 kDa at 31.8 minutes and the non-phosphorylated FHA-2 domain had a calculated mass of 10.940 kDa at 33.5 minutes.

6.8 A hypothetical model of how PknF could be controlling Rv1747 function in *M. tuberculosis*

Through combining all the data generated in this study a schematic representation of a potential mechanism by which PknF controls Rv1747 function was proposed and is diagrammatically shown in figure 6.26. Upon sensing an unknown signal PknF becomes autophosphorylated within the activation loop on residues T173 and T175 rendering the kinase functionally active. The peptide binding pocket of either of the Rv1747 FHA domains may then bind to one of the activation loop threonine residues. The function of the FHA-1 domain has been shown to be essential for Rv1747 function. Binding of the FHA domain also delivers the substrate protein into the active site of the kinase enabling substrate phosphorylation on residues T150 and T208. Upon substrate release, either of the FHA domains can then recognise and bind to either of the phosphothreonine epitopes on Rv1747, causing a conformational change within the protein structure to render it into an active conformation for the transport of the substrate across the *M. tuberculosis* cell membrane which is necessary for a virulent infection.

6.9 Discussion

6.9.1 *ΔRv1747* is attenuated for growth in BMDMs

The bacterial load of the *ΔRv1747* mutant in BMDM infections was significantly impaired in both naïve and IFNγ activated cells compared to its parental wild type strain (figure 6.2). Furthermore, the mutant phenotype was lost when the *Rv1747* gene was replaced in the chromosome confirming that the growth attenuation observed was due to the deletion of *Rv1747*. The results obtained demonstrated that the *Rv1747* deletion strain is still attenuated for growth in murine BMDM infections and supports the hypothesis that Rv1747 is indeed required for a virulent infection. Furthermore, these results corroborate an earlier study where the *Rv1747* deletion strain was also shown to be attenuated for growth (Curry *et al.*, 2005), and demonstrates that this mutant phenotype has been stably inherited over many generations of growth.

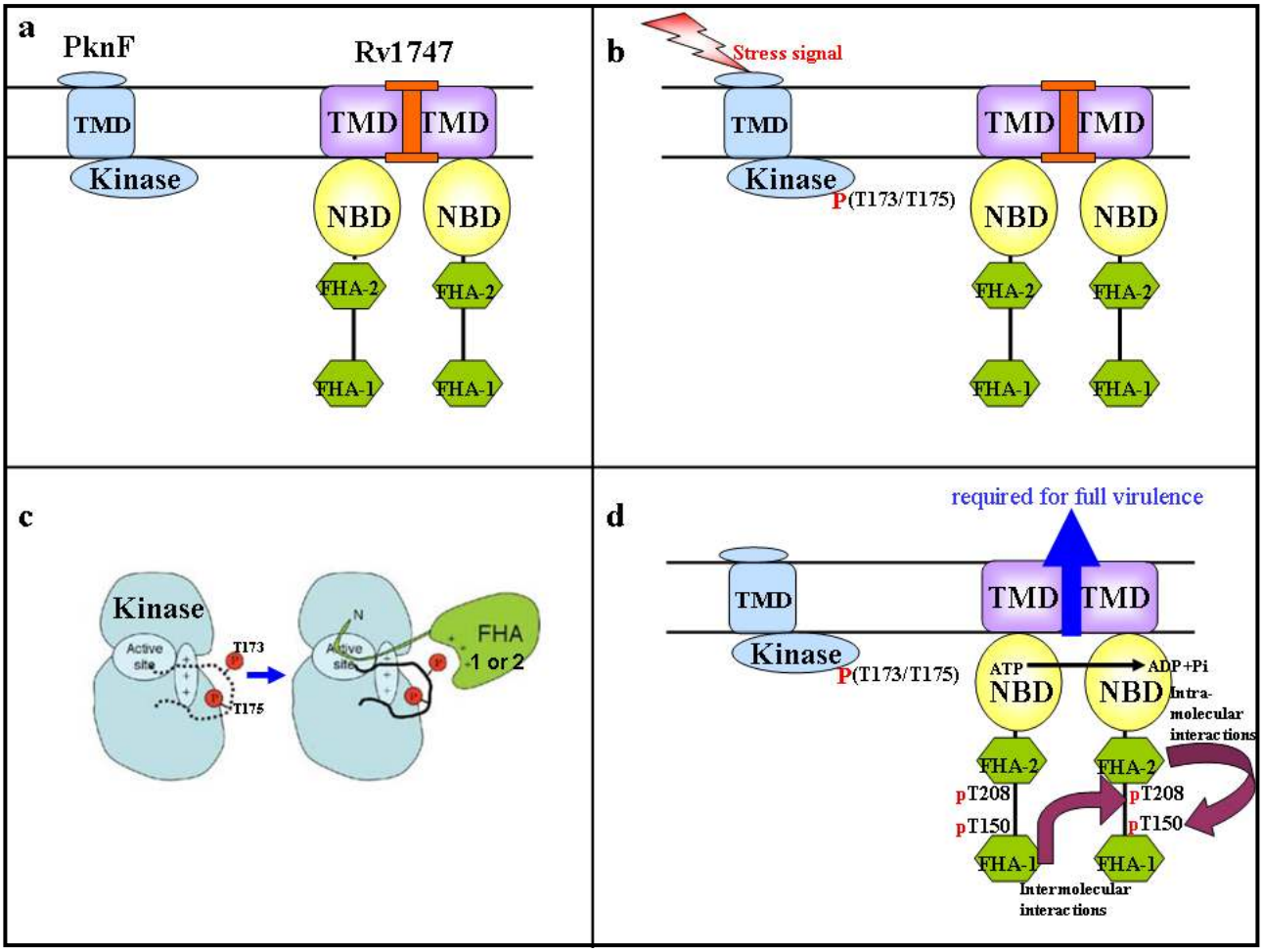


Figure 6.26. Schematic representation of a potential mechanism by which PknF controls Rv1747 function in *M. tuberculosis*. In the absence of Rv1747 phosphorylation the ABC transporter is inactive (figure 6.26a). Upon sensing a stress signal PknF autophosphorylates on residues T173 and T175 of the activation loop (figure 6.26b). The peptide binding pocket of the FHA-1 or FHA-2 domain is then able to bind one of the PknF activation loop threonine residues and deliver the substrate into the active site for phosphorylation on residues T150 and T208 (figure 6.26c - adapted from Villarino *et al*., 2005). Through presumed intra- and inter-molecular interactions between the FHA domains and the pT residues of both halves of the transporter the function of Rv1747 is activated and the substrate can be exported (figure 6.26d). Energy for transport is driven by the hydrolysis of ATP by the nucleotide binding domains (figure 6.26d). Protein domains are not drawn to scale. The Rv1747 transmembrane domains have been simplified and the two black lines in figures 6.26a, b and d represent an oversimplification of the *M. tuberculosis* cell wall. Note in figure 6.26d that the substrate could be transported to another part of the cell wall and not necessarily exported from the cell.

IFNγ activation of the BMDMs resulted in reduced growth of all the *M. tuberculosis* strains compared to the growth seen in the naïve cells. In the IFNγ activated cells it can be noted that fewer cells of the *ΔRv1747* strain were phagocytosed inside the macrophages at 6 hours post infection compared with the wild type and complementing strains (figure 6.2b). The same inocula were used to infect both the naïve and IFNγ activated cells. In naïve cells the number of *M. tuberculosis* cells counted at 6 hours post infection was very similar for all three strains (figure 6.2a). Nonetheless, if the IFNγ data in figure 6.2b are plotted normalised to the initial 6 hour uptake the *ΔRv1747* strain is still significantly attenuated for growth (data not shown). However, in figure 6.3b it can be clearly seen that the uptake of the strains into both naïve and IFNγ activated macrophages in this infection was very similar and thus the most likely explanation for the lower uptake observed in figure 6.2b was due to the variability in the data obtained in macrophage infections in general. Although patterns of data are consistent in BMDM infections i.e. a mutant with growth attenuation can always be observed, the individual time points can vary quite considerably from experiment to experiment. This is probably due to a number of reasons; mouse age (typically varies from 6 to 8 weeks), variability in uptake of the *M. tuberculosis* cells between the wells of the three technical replicates and variability in the serial dilutions of the lysed macrophages containing the *M. tuberculosis*.

The Griess nitrite assay showed that cells treated with IFNγ were indeed activated as demonstrated by the presence of NO/nitrite in the supernatants of infected BMDMs (figure 6.1). The results also showed that there were no differences in the levels of nitrite produced when IFNγ stimulated cells were infected with wild type, *ΔRv1747* and *Rv1747* complement strains, suggesting that the reason the *ΔRv1747* deletion strain had a lower bacterial burden was not due to an increase in the nitrite concentration in the supernatant of cells infected only with this mutant strain which in turn may have lead to slower growth or an increased killing of this strain.

6.9.1.1 Assessment of the growth of the threonine mutant strains in BMDMs and mice and implications for the regulation of the PknF-Rv1747 system

6.9.1.1.1 BMDM infection

The two phospho-threonine sites identified in this study (figure 5.6), which were demonstrated to be important for the *in vitro* phosphorylation of Rv1747 by PknF (figure 5.8), were mutated to either warrant them incapable of receiving a phosphate group (T150A/T208A) or to mimic constitutively phosphorylated residues (T150D/T208D) within the *Rv1747* complementing strain and tested for their ability to rescue the growth attenuation phenotype of the *ΔRv1747* strain in BMDM infection experiments. It was demonstrated that the bacterial load of both the T150A/T208A and T150D/T208D strains were attenuated for growth at 120 and 168 hours post infection in naïve cells and also at 168 hours post infection in IFNγ activated cells (figure 6.3). However, the growth attenuation of both these strains was intermediate between the *ΔRv1747* deletion and wild type strain in IFNγ activated cells whereas in naïve cells both the point mutation strains appeared to grow similar to the *ΔRv1747* deletion strain.

Aspartic acid is a negatively charged amino acid whose structure is very similar to that of a serine or threonine residue which has been phosphorylated. Therefore, this amino acid, along with glutamic acid which has similar properties to aspartic acid, are both used in phosphorylation studies to act as phosphorylation mimics (Veyron-Churlet *et al*., 2009). However, one problem with the construction and use of these mimics is that they do not necessarily always act as phosphorylation mimics inside the cell. In this study the T150D/T208D phosphorylation mimic strain did not rescue the growth phenotype of the *ΔRv1747* strain; furthermore, the T150A/T208A phosphorylation null strain also did not rescue the growth attenuation phenotype of the *Rv1747* deletion. Therefore it can be hypothesised that perhaps the phosphorylation mimic strain did not mimic phosphorylation successfully in terms of protein function; perhaps the FHA domains cannot bind aspartic acid residues and this is why the growth attenuation phenotype was not restored in this strain. Furthermore because the T150A/T208A strain also did not rescue the growth attenuation phenotype then this suggests that **phosphorylation by PknF positively regulates the activity of the Rv1747 ABC transporter** i.e. phosphorylation of the

T150 and T208 residues of Rv1747 by PknF serves to transduce the signal which activates the function of the transporter. This hypothesis corroborates the microarray data which showed that the same differential changes in gene expression were noted in both of the mutant strains, for example, the *iniBAC* operon was upregulated in both of the deletion strains (chapter 4).

6.9.1.1.2 Murine infection

The mouse aerosol infection experiment demonstrated that the *ΔRv1747* strain was attenuated for growth both in the lungs, and more significantly so in the spleens (figure 6.4); earlier findings from an intravenous infection model found that the *Rv1747* deletion strain was attenuated for growth by a maximum of ten fold (1 log) in the lungs and spleens compared with wild type throughout the time course of infection (Curry *et al.*, 2005). Results from this current study showed the growth phenotype of the *ΔRv1747* mutant to be more pronounced; in the lungs the mutant was attenuated by a maximum of 2.5 logs at 30 days post infection and in the spleens the mutant was attenuated for growth by 2.8, 3, and 3.8 logs at 30, 60 and 90 days respectively. Compared with the intravenous route of infection, an aerosol infection model is a better way to study *M. tuberculosis* virulence and pathogenesis because mice are infected through inhalation of bacilli, mimicking the mode of infection in humans, whereas in an intravenous infection the mice have TB bacilli injected into their tail vein; in this type of experiment approximately 10% of the inoculated bacteria travel straight to the spleens of the animals and only 1% goes to the lungs (J. Dillury, Personal Communication). This route is in effect the opposite way round to the mode of infection in humans. Therefore it was important in this study to assess the growth of the *ΔRv1747* deletion, complement, and both the mutated complement strains in the aerosol infection model. In a virulent *M. tuberculosis* infection, from the lungs the bacteria first spread to the lymph nodes and then to other organs including the spleen and liver via the circulatory system.

Interestingly, at day 30, mice infected with the *ΔRv1747* and *Rv1747* complement T150A/T208A strains had no *M. tuberculosis* bacilli present in their spleens (figure 6.4b). Furthermore, at day 60 only three of the five animals infected with the *ΔRv1747* strain had bacilli in the spleens suggesting that both this mutant and the *Rv1747* complement T150A/T208A strain are highly attenuated for growth in the

spleens, highlighting the critically important role that the Rv1747 ABC transporter plays in *M. tuberculosis* pathogenesis. These results therefore corroborate the macrophage data again suggesting that **PknF positively regulates the activity of the Rv1747 ABC transporter**. Furthermore, the *Rv1747* complement T150D/T208D strain also did not fully complement the *Rv1747* deletion growth phenotype in the lungs or spleens and the most logical explanation is that this strain is not successfully acting as a phosphorylation mimic. Importantly, FHA domains cannot recognise aspartate residues (S. Smerdon, Personal Communication); therefore in a T150D/T208D mutant there can be no intra- or inter- molecular interactions between the FHA domains and these residues explaining why this mutant had a similar phenotype to the alanine mutant. Clearly, one also has to bear in mind that there may be another phosphorylation site present in Rv1747 that was not identified in the LC/MS/MS screen; indeed, in the *in vitro* phosphorylation assays there was a very faint band present corresponding to the size of Rv1747 in the T150A/T208A mutant (figure 5.8). If this were the case it may offer an explanation as to why neither of the double mutated complements displayed a growth phenotype as severe as the *ΔRv1747* strain. Furthermore, the *Rv1747* complement strain did not fully complement the *ΔRv1747* growth phenotype in the lungs or especially in the spleens (figure 6.4). The complement contains an extra copy of *pknF* (figure 1.12) which may have regulatory effects on the function of Rv1747 offering an explanation why the growth attenuation was not fully complemented.

One hypothesis to discuss is whether the *ΔRv1747* and *Rv1747* T150A/T208A complement strains (and to some extent the T150D/T208D strain) are defective in dissemination from the lungs to the spleens or whether these strains just harbour a growth defect *in vivo*. A 2001 paper described a heparin-binding haemagglutinin protein of *M. tuberculosis* (*hbhA*, *Rv0475*) which was shown to be required for extra-pulmonary dissemination. A deletion mutant of this protein grew as well as wild type in the lungs and spleens of mice infected intravenously (Pethe *et al*., 2001). However, in an aerosol model of infection the growth of this mutant was indistinguishable from that of the wild type in the lungs but was severely attenuated for growth in the spleens (Pethe *et al*., 2001). These data conclusively showed that the mutant harboured a defect in dissemination because the strain grew at the same rate as the wild type in the spleens of mice infected via the intravenous route. In this

current investigation *ΔRv1747* was not only attenuated for growth in the lungs and spleens of mice infected via the aerosol route, but was also attenuated in both organs in an intravenous infection (Curry *et al*., 2005). Taken together, this suggests that the *ΔRv1747* mutant strain is defective in growth in general *in vivo* and therefore is not required for extra-pulmonary dissemination.

However, another study published in 2004 stated that an *lspA* mutant was also defective in extra-pulmonary dissemination (Sander *et al.*, 2004). Their study showed that deletion of a lipoprotein signal peptidase (*lspA*, *Rv1539*) in *M. tuberculosis* did not affect *in vitro* growth but did result in significant growth attenuation in mouse macrophages and in the lungs of BALB/c mice following an aerosol route of infection; critically, like *ΔRv1747*, this mutant failed to colonise the spleens of the infected animals over the time course of the experiment (42 days) (Sander *et al.*, 2004). Full complementation of growth was observed when the *lspA* gene was replaced into the chromosome and the mycolic acid pattern was similar for all three strains (Sander *et al.*, 2004). The authors stated that these results showed that the *lspA* gene was required for dissemination of infection from the lungs to the spleens, even though the mutant strain was also attenuated for growth in the lungs.

One could hypothesise at this stage that perhaps Rv1747 functions to export one of the 99 *M. tuberculosis* lipoproteins to the cell surface which is critical for pathogenesis because the *in vivo* phenotypes observed with the *Rv1747* and *lspA* deletion strains were similar (Kovacs-Simon *et al.*, 2010). To investigate this one could prepare cell extracts from wild type, *ΔRv1747* and the *R1747* complementing strain and probe for lipoproteins using specific antibodies.

6.9.1.1.3 BMDM infections with the single threonine mutants

Assessment of the panel of single threonine mutants in the BMDM infection model showed that all four single mutant strains had a bacterial load similar to that of the wild type and the *Rv1747* complementing strain throughout the time course of infection, with the exception of the T150D strain which displayed a growth attenuation phenotype 120 hours post infection (figure 6.5). However, by 168 hours post infection there was no longer a significant difference between the growth of this strain and the wild type strain. The data suggests that the reason why a phenotype

was observed at 120 hours was due to the inherent variability seen in macrophage infection experiments.

6.9.1.1.4 Infection summary

In summary, it was shown that the T150A/T208A and T150D/T208D strains displayed a growth attenuation phenotype in BMDMs and mice similar to that of the *ΔRv1747* strain; the most logical explanation of the results was that the T150A/T208A mutation did not rescue the phenotype of the mutant strain because phosphorylation of Rv1747 by PknF positively regulates the activity of the transporter, and secondly, the phosphorylation mimic strain failed to render the Rv1747 transporter constitutively active because the FHA domains are unable to interact with aspartate residues. Furthermore, the fact that none of the single mutants displayed a marked growth phenotype suggested that the ABC transporter is still able to function to a degree even if one of the threonine residues is mutated and therefore mutation of both the T150 and T208 residues are necessary to prevent Rv1747 transporter function. Perhaps the ABC transporter can function to a degree to rescue the growth attenuation phenotype if either one of the threonine residues is mutated but not if both residues are rendered unable to be a phospho-acceptor. These data also suggest a degree of redundancy between the two residues.

6.9.1.2 Cytokine assays

Cytokine ELISAs demonstrated there were no differences in the production of IL12-p40 throughout the course of infection at two different MOIs (figure 6.6). When examining TNFα production there were no significant differences in the concentrations of secreted cytokine produced when the MOI was 5:1 (figure 6.7). However, when the MOI was 0.5:1 at the 6 hour time point in IFNγ activated macrophages infected with the wild type strain the cells secreted 19 pg/ml of TNFα whereas macrophages infected with the mutant and complement strains produced levels below that of the negative control well. Furthermore, at 120 hours post infection the IFNγ activated cells secreted more TNFα when infected with the wild type and *ΔRv1747* strains compared to the *Rv1747* complementing strain. The reason for this is unclear as at an MOI of 5:1 no differences were noted. The sensitivity of the assay may offer an explanation of the result at 6 hours post infection as a

concentration of 19 pg/ml is very low so for the other two strains the levels of TNFα could easily have been undetectable. The results of IL6 secretion into the culture supernatants after infection with all three *M. tuberculosis* strains at an MOI of 0.5:1 showed that at all time points in both naïve and IFNγ activated macrophages there were no significant differences in the levels of IL6 secreted between the wild type and *ΔRv1747* strains (figure 6.8). However, at 6 and 72 hours post infection in the IFNγ activated cells, lower levels of IL6 production were observed from the macrophages infected with the *Rv1747* complement strain compared to the wild type and deletion strains. The reason for this is unclear but what is important is that the *Rv1747* complementing strain was able to fully rescue the phenotype of the *ΔRv1747* deletion strain. The results are from three technical replicates of one biological replicate so the assay should be performed again before final conclusions are drawn. **The important conclusion from these assays was that there were no significant differences between the wild type and mutant strains which could have accounted for the growth attenuation phenotype due to a difference in cytokine response.**

One major limitation of this experiment is that the cytokine response which can be examined in infected macrophages over seven days does not mimic the response seen in infected animals over many months; perhaps a better experiment to perform would be to look at the cytokine response in infected mice.

It was hypothesised in this study that the *ΔRv1747* mutant may have induced a different cytokine response due to the transporter not performing its function and this could be a reason why this strain fails to replicate/survive as well in macrophages; however, this was not the case. Although no major differences were noted in this investigation, previous studies have characterised mutants that induce different cytokine responses and have been able to relate those differences to the function of the protein. A study examining the effects of the deletion of the *pknE* gene found that infection of THP-1 macrophages with this mutant caused the THP-1 cells to secrete significantly lower levels of the pro-inflammatory cytokines TNFα and IL6 in the culture supernatants compared to the wild type and complementing strains (Jayakumar *et al.*, 2008). Their study also showed that the *pknE* promoter sensed nitric oxide stress and furthermore they demonstrated that addition of nitric oxide

donors to macrophages induced increased apoptosis in those cells infected with the *ΔpknE* strain which suggested that PknE played a role in the prevention of apoptosis. They then related these findings to propose a mechanism whereby PknE senses nitric oxide stress and then enhances macrophage viability by preventing apoptosis by interfering with the host signalling pathways (demonstrated by the reduced cytokine levels in the *ΔpknE* mutant compared with the wild type and complementing strains) (Jayakumar *et al.*, 2008). Other studies on *M. tuberculosis* mutants have also found differences in the cytokine response including a study on the *mce1* operon in which the mutant strain produced less TNFα and IL6 compared to the parental strain in a murine macrophage infection experiment (Shimono *et al.*, 2003). This strain was also found to be hypervirulent in mice as it was unable to enter a stable persistent phase in the lungs of mice and all mice were dead by week 41 of the experiment compared with the wild type control where all the mice were still alive (Shimono *et al.*, 2003). A 2007 study found that the *Rv3083-Rv3089* operon of *M. tuberculosis* was induced 17-33 fold when exposed *in vitro* to acidic conditions (Cheruvu *et al.*, 2007). Furthermore the mutant displayed a growth attenuation phenotype in cell lines and THP-1 cells infected with this strain produced less IL1ß, IL6 and IL8 compared with the wild type strain in the first 48 hours of infection (Cheruvu *et al.*, 2007). Taken together, these studies highlight the importance of investigating the cytokine responses to *M. tuberculosis* infection.

Interestingly, no differences in cytokines were observed between the wild type and *ΔRv1747* strains even though the numbers of intracellular bacteria were significantly different by 120 and 168 hours post infection (figure 6.2). Assessment of IL12-p40 and TNFα levels were stopped at 120 hours post infection and assessment of IL6 levels were only performed for the first 72 hours of infection; perhaps these studies should be extended to analyse the later time points.

6.9.2 PknF

It was unknown whether the *pknF* gene played a role in the pathogenesis of *M. tuberculosis*. It was hypothesised that by examining the growth of this mutant and its respective complementing strain in BMDMs it may also provide further evidence as to whether this kinase positively or negatively regulates the activity of the Rv1747 ABC transporter.

6.9.2.1 BMDM infections

Results showed that in naïve BMDMs there was no growth phenotype of the *ΔpknF* null or complement strains up until 168 hours post infection when a significant difference was observed in the bacterial load of the mutant and complement strains compared with the wild type strain (figure 6.9). In IFNγ activated BMDMs the growth of all three strains were comparable up until 168 hours post infection where the *ΔpknF* strain had a significantly lower bacterial load compared with the wild type and *pknF* complementing strains. **The results therefore again suggested that PknF phosphorylation of Rv1747 positively regulates the activity of the transporter because when *pknF* was deleted from the chromosome this resulted in a growth attenuation phenotype.**

However, it was surprising that the *pknF* complementing strain had a significantly different bacterial load to that of the wild type strain in naïve cells at 168 hours post infection. In some previous *M. tuberculosis* studies it has not been possible to rescue the effects of a gene deletion by simple replacement of that gene back into the chromosome. However, in the case of this study the *pknF* complement did rescue the differential gene changes observed in the microarray and qRT-PCR studies suggesting that the function of the kinase is restored in this strain (see chapter 4). Results also showed that there was a difference in the 6 hour bacterial load uptake values between the three strains; in both naïve and IFNγ stimulated cells the wild type strain had a greater bacterial load compared to the mutant and complementing strains (figure 6.9). The reason for this result could be due to an error in the dilution of the inocula or due to a phenotype of the *ΔpknF* deletion and complementing strains themselves. The results were then normalised to the initial six hour uptake values of each strain; normalised values showed that the *pknF* complement has a

similar pattern to the wild type strain; furthermore, the *ΔpknF* strain followed a similar pattern to 120 hours post infection and then became attenuated for growth between 120 and 168 hours post infection (figure 6.10). However, this type of analysis should never be taken at face value because a real phenotype of the *pknF* deletion strain may be that it is not phagocytosed as efficiently as the wild type strain. In addition, if the experiment had been terminated at 120 hours post infection then no growth attenuation phenotype of the *ΔpknF* strain would have been observed regardless of how the data was analysed.

Furthermore, the results showed very large error bars for the *pknF* complement strain in IFNγ activated cells. This was because between the three technical replicates the number of colonies obtained varied quite substantially. This was also observed in the second biological replicate of this experiment. Perhaps there is an additional mutation within this strain that is accounting for this phenotype; to investigate this the *pknF* complementing strain could be sequenced.

One possible reason why a strong growth attenuation phenotype was not observed with the *ΔpknF* strain in BMDM infections is that there are 10 other kinases in the *M. tuberculosis* genome which may be able to phosphorylate Rv1747 to activate the function of the transporter. Indeed, by *in vitro* phosphorylation assays domains of Rv1747 have been shown to be phosphorylated by PknE, PknD, PknB and PknF (Grundner *et al.*, 2005).

To test the *ΔpknF* strain more thoroughly and to determine whether this kinase is required for the growth of *M. tuberculosis in vivo* the *pknF* deletion strain should now be tested in a murine infection model to determine if this strain is attenuated for growth in both the lungs and spleens of mice. This would perhaps give a clearer indication of the requirement for this kinase for the *in vivo* growth of *M. tuberculosis*.

6.9.3 Characterisation of the function of the Rv1747 FHA domains

6.9.3.1 BMDM infections

6.9.3.1.1 FHA domain function requires specific conserved residues

Highly conserved residues within FHA domains directly contact the phosphopeptide, either by contacting the peptide backbone or by forming a network of hydrogen bonds with the pT residue on the epitope (Durocher & Jackson, 2002). In the well studied Rad53 FHA domain-containing protein, there are six highly conserved residues in the FHA-1 domain, five of which are located around the peptide binding site. Residues arginine-70 and serine-85 bind directly to the pT residue itself (Durocher *et al.*, 2000). Consequently, mutation of the conserved serine 85 residue within one of the peptide binding loops abolishes the binding activity of the FHA domain to the pT residue. The orthologous serine residue in Rv1747 FHA-1 is serine 47 and in the FHA-2 module is serine 248. Data in this study has shown that mutation of the serine 47 residue in the FHA-1 module resulted in severe attenuation of growth in BMDM infection experiments in both naïve and IFNγ activated cells whereas the FHA-2 domain S248A mutant demonstrated no growth phenotype compared with the wild type strain (figure 6.11).

6.9.3.1.2 Conclusions drawn from Rv1747 FHA domain BMDM infections

A number of conclusions can be drawn from the data. As the FHA-1 S47A mutation resulted in a growth attenuation phenotype it **again suggests that regulation of Rv1747 by PknF is positive because when the FHA-1 domain can no longer bind a pT epitope either in the PknF activation loop or residues T150 or T208 of Rv1747 it resulted in a growth phenotype similar to that observed in the *Rv1747* deletion mutant** (figure 6.2) suggesting that the FHA-1 mutation renders Rv1747 unable to function. These results again support the earlier microarray and qRT-PCR data where the same pattern of differential gene regulation was observed in both the kinase and transporter mutant strains suggesting that there was a positive mechanism of regulation between PknF and Rv1747 (chapter 4).

Furthermore, the growth attenuation data gathered for the FHA-1 S47A mutant suggests that the function of the FHA-2 domain alone within this mutant strain is not

sufficient for activation of the transporter activity of Rv1747. This makes sense because out of the six FHA domain containing proteins in *M. tuberculosis*, Rv1747 is the one to contain two FHA modules and there must, therefore be a functional reason for this. Perhaps the FHA-2 domain functions in intramolecular associations with the T150 and/or T208 residues on Rv1747 and it is the FHA-1 domain which binds to the PknF activation loop to allow substrate phosphorylation. If indeed the case, this hypothesis may offer an explanation as to why the FHA-1 domain is critical for Rv1747 protein function whereas from the data generated in this study the FHA-2 domain does not appear to be as important.

Mutation of the S248 residue in FHA-2 will also abolish the capability of this FHA domain to bind to pT residues; however, from the data generated it appears that the presence of a functional FHA-1 domain in this strain was able to rescue the growth attenuation phenotype, i.e. the ABC transporter is able to function enough so that the growth phenotype seen in the BMDM infections is rescued. As we have no other readout of phenotype for the *ΔRv1747* strain apart from growth attenuation in BMDM and mouse infection experiments we are unable to speculate any further at this stage about the ability of the transporter to function in the FHA-2 S248A domain mutant.

Alternatively it could also be argued that it is the FHA-2 domain which functions to bind the activation loop of PknF and then the FHA-1 domain functions in intramolecular associations with T150 and T208 residues causing a conformational change to activate the activity of the transporter. However, this seems unlikely due to the fact that in the FHA-2 S248A mutant no growth phenotype is observed suggesting that in this mutant Rv1747 is phosphorylated by PknF and is functional to some extent. A third mechanism of regulation could involve both the FHA domains binding the activation loop of PknF; perhaps the binding of one FHA domain results in the phosphorylation of the T150 residue and binding of the other FHA module results in the phosphorylation of the T208 residue and then intramolecular associations happen if and when required to regulate the transporter activity of Rv1747. Unfortunately from the data generated in this study we are unable to speculate further at this time as to the precise mechanism of regulation.

Interestingly in the earlier 2005 study examining the interaction between PknF and Rv1747 in a yeast 2-hybrid system, it was found that the interaction was completely abrogated in an S47A mutant and was reduced by approximately three fold in a S248A mutant corroborating the evidence founding this study that the FHA-1 domain is critically important for Rv1747 function (Curry *et al.*, 2005).

Taken together, these results suggest that **the FHA-1 domain is critically important for the function of the Rv1747 transporter**. To further investigate the importance of the Rv1747 FHA domains for *M. tuberculosis* pathogenesis animal infection studies with these mutant strains should be performed. Furthermore, a strain should be created in which both the Rv1747 FHA domains are mutated at their critical serine residues in the peptide binding loop; it can be hypothesised that this should result in the same growth attenuation phenotype as observed in the *ΔRv1747* and FHA-1 S47A mutants.

6.9.3.2 Determination of the binding kinetics between FHA domains and pT peptides

FHA modules are phosphothreonine epitope binding domains and the recruitment of the binding domain to the epitope results in transmission of the signal; in this case the binding of the Rv1747 FHA domains to the pT epitopes are presumed to control the activity of the ABC transporter.

6.9.3.2.1 Use of ITC to determine binding kinetics

ITC demonstrated that both FHA domains bound to both the Rv1747 phosphopeptides with similar binding affinities and furthermore showed that the data generated was indeed due to the direct binding of the FHA domains to the phosphopeptides because in an FHA-1 S47A mutant all binding was abolished. As already discussed, in the Rad53 FHA-1 domain residues arginine-70 and serine-85 bind directly to the pT residue itself (Durocher *et al.*, 2000). Consequently, mutation of the conserved serine 85 residue within one of the peptide binding loops abolishes the binding activity of the FHA domain to the pT residue. The homologous serine residue in Rv1747 FHA-1 is serine 47 and this study has already demonstrated that this serine residue is critical for the function of the Rv1747 transporter (figure 6.11).

Since the data showed that both FHA domains bound both the phospho-threonine peptides with similar affinities it suggested that there was not a preferred peptide for each of the FHA domains. However, one has to bear in mind that *in vivo* both of the pT residues may not be available to both of the FHA domains for binding since physical constraints may prevent this from happening; for example the T208 site is very close to the FHA-2 domain so it would be not possible for this domain to rotate in such a way to bind the epitope unless the domain from one of the 'half-transporters' bound the epitope on the other polypeptide. This hypothesis is discussed in the subsequent section.

Furthermore, the fact that both the Rv1747 FHA domains bound both the Rv1747 pT peptides is not surprising; it is probably the result of epitope co-evolution between the FHA domains and the serine/threonine protein kinases. As previously discussed serine/threonine protein kinases and FHA domains have probably co-evolved to recognise overlapping consensus motifs.

The binding kinetics between the Rv1747 FHA domains and non-Rv1747 peptides were also determined. Data demonstrated that FHA-1 and FHA-2 domains could also bind the pT22 peptide from Rv1827 with dissociation constants in the μM range (22 μM and 21 μM respectively) and furthermore FHA-2 could also bind a phospho-peptide from Rv0020c with a K_d of 9 μM. The ability of the FHA-1 domain to bind this peptide was not assessed. Titrations of the Rv1827 FHA domain with the pT22 peptide resulted in a K_d of 12.8 μM, similar to what was determined for the Rv1747 FHA domains (T. Nott, personal communication).

6.9.3.2.2 FHA domain pT-containing epitope specificity

For FHA domains to bind pT-containing epitopes the residue at position pT+3 has been shown to be critically important for this interaction (Durocher *et al.*, 1999). The Rad53 FHA-1 phosphopeptide interaction was able to tolerate amino acid substitutions at pT-3 to pT-1and pT+1 to T+2 but the interaction was abolished when the pT+3 residue was mutated highlighting the importance of this residue for the interaction and suggesting that this residue is involved in the selectivity of binding (Durocher *et al.*, 1999). Studies in 2000 confirmed this binding selectivity theory

using anchored peptide libraries to identify FHA domain consensus binding motifs (Durocher *et al.*, 2000; Liao *et al.*, 2000).

The Rv1747 pT150 and pT208 phosphorylation sites contain hydrophobic residues in the pT+3 position, namely isoleucine for pT150 and methionine for pT208 as does the pT22 peptide from R1827 which has a phenylalanine residue at pT+3. The presence of a hydrophobic residue could explain why the Rv1747 FHA domains could bind all three peptides although FHA-2 was also shown to bind the pT377 peptide from Rv0020c which contains a non-hydrophobic threonine residue at the pT+3 position. However, amino acids I, M, F, and T were all present at the pT+3 position in the phosphorylation site motif analysis for PknF suggesting that through co-evolution the Rv1747 FHA domains could also potentially recognise pT containing epitopes with any of these residues at the pT+3 position (Prisic *et al.*, 2010). For PknB, which controls Rv1827 and Rv0020c by phosphorylation, residues overrepresented at the pT+3 position included I, F and M highlighting the overlapping consensus motifs of the two kinases (Prisic *et al.*, 2010).

Inside the *M. tuberculosis* cell there are spatial and temporal separations of protein expression and the bacterium will have evolved to minimise the necessary complexity. Therefore again it was not unexpected that the Rv1747 FHA domains could also bind to the Rv1827 and Rv0020c phosphopeptides as *in vivo* an FHA domain on Rv1747 is not likely to come into contact with a phosphopeptide on Rv1827 or Rv0020c.

6.9.3.2.3 Principles of ITC and further explanation of results

In ITC the signal that is generated is the heat of binding. Binding of the protein to a peptide can be an exothermic or endothermic reaction (due to the net outcome of hydrogen bond formation and electrostatic interactions); furthermore, as the FHA domain binds the peptide some buffer has to move out of the binding pocket and all these occurrences cause temperature fluctuations. In this study binding of the FHA-1 and FHA-2 domains to phosphopeptides were always exothermic reactions. This is demonstrated by the series of heat spikes in the top half of each figure which plotted the amount of power (in μcal/sec) needed to maintain the reference and sample cell at the same temperature; as each reaction was exothermic it means the instrument

needed to use less energy to heat the cell which resulted in the generation of negative heat spikes. As each experiment progressed the binding sites on the peptide were gradually saturated and therefore the heat effect was reduced. This can be observed in the top panel of figures 6.13, 6.14, 6.16 and 6.17. Intriguingly, binding of the FHA-2 domain to phosphopeptides consistently gave a stronger heat signal upon each injection of protein into the cell compared to the FHA-1 domain. The heat spikes generated in the FHA-1 S47A studies were due to the heat of dilution that occurs when two protein species are mixed, this occurrence is not associated with protein binding.

The lower panel in each figure showed the total heat effect per injection by integrating the area of the peaks and plotting them against the molar ratio of protein and ligand. A curve was then fitted and this generated the binding data. Several iterations of the fit of the curve were performed in order to get the best fit and the Chi-squared value gave information about the fit of the curve to the data. From the shape of the curve it was possible to determine the stochiometry and enthalpy of the reaction as well as the binding affinity (K), the inverse of which is the dissociation constant; $K_d = 1/K$. The binding stoichiometry (N) of the interactions investigated in this study should have been 1:1 as one FHA molecule can bind only one peptide molecule. Results showed that for FHA-1 interactions with Rv1747 peptides N was always approximately 1 or just above (figure 6.13). This shows that the calculated protein concentration of the FHA domain was approximately correct. For the FHA-2 titrations with Rv1747 peptides N was 0.74 and 0.79 respectively (figure 6.14a and b). When N was less than 1 it suggested that the protein concentration was less than what was determined; the FHA-2 construct does not contain tryptophan, tyrosine or phenylalanine residues which contain conjugated aromatic rings enabling them to absorb light in the ultraviolet range. Consequently protein concentration could not be calculated by spectroscopy and had to be determined using the refractive index of the protein by SEC-MALLS; the ITC results demonstrated that the refractive index was probably not very accurate hence why the N ended up being less than 1. N was also slightly less than 1 in the titrations between both Rv1747 FHA domains and the non-Rv1747 peptides with values of 0.84, 0.82 and 0.84 respectively reflecting the high degree of precision of the instrument (figures 6.16 and 6.17).

The enthalpy change (ΔH) in all titrations was negative due to the gaining of new bonds when binding occurred resulting in a release of heat (exothermic reaction) (figures 6.13, 6.14, 6.16 and 6.17). In this study an increase in disorder (entropy (ΔS)) occurred when titrating FHA-1 against both the Rv1747 phosphopeptides (figure 6.13), when titrating FHA-1 against pT22 from Rv1827 (figure 6.16a) and when titrating FHA-2 against the pT377 peptide from Rv0020c (figure 6.17). However, when titrating FHA-2 against the Rv1747 peptides (figure 6.14) or the Rv1827 peptide (figure 6.16b) the ΔS was negative. This was an intriguing result as it suggested that the system was becoming more ordered as protein binding occurred. When the ΔS was negative it can be noted that there was also a stronger negative ΔH of binding compared with when the ΔS was positive, presumably to compensate for the negative ΔS to allow binding to occur (see figures 6.13 and 6.14). Based on the theory of energy changes given by the equation $\Delta G = \Delta H - T\Delta S$ where G is free energy content, H is enthalpy, T is temperature in Kelvin and S is entropy, for reactions to be favourable the sum of the free energy changes must be negative. If ΔS is negative then there must be a large negative ΔH of binding to make a negative ΔG to allow binding to occur.

In summary, ITC demonstrated that **both the FHA domains could recognise and bind to both the identified pT residues on Rv1747 lending further support to the hypothesis that the identified residues do indeed play an important role in the *pknF-Rv1747* signalling system**.

6.9.3.3 Does the FHA-2 domain dimerise upon phosphorylation at T208?

It was hypothesised that when residue T208 on Rv1747 was phosphorylated by PknF then the FHA-2 domain on one half of the transporter may bind the T208 residue on the other half and vice versa in order to transduce the signal and regulate the activity of the Rv1747 transporter. It was demonstrated that PknF could phosphorylate FHA-$2^{202\text{-}310}$ *in vitro*. The amount of radioactivity incorporated into the FHA-2 construct after phosphorylation by PknF was much greater that the amount incorporated by the FHA-$1^{1\text{-}120}$ domain construct (figure 6.19). This is probably because T208 is present within the FHA-2 construct whereas for the FHA-1 construct no pT residue was identified between residues 1 and 120 although the data does suggest that FHA-$1^{1\text{-}120}$ is also phosphorylated to some extent due to the incorporation of radioactivity.

Using electrospray mass spectroscopy it was shown that after four hours of incubation of $PknF^{1-292}$ with FHA-$2^{202-310}$ and ATP almost all the FHA-2 protein was phosphorylated on one residue and a small proportion of the protein construct contained two phosphorylated residues demonstrated by the increase in mass of 160 Da. Importantly, none of the protein was in its non-phosphorylated state. Electrospray mass spectrometry gives the precise molecular weight/s of the species in the protein sample but cannot inform on which residue the addition of a phosphate group occurred. For this investigation it had to be assumed that the phosphorylated residue was at position T208 because this was one of only two phosphorylated threonine residues identified that were phosphorylated by PknF; however, there is no conclusive evidence to prove this. Furthermore, there was a small abundance of the protein that contained two phosphorylated residues and it had to be hoped that this would not interfere with the experiments to determine if FHA-$2^{202-310}$ became dimeric after phosphorylation. For future investigations LC-MS/MS could be employed to determine the residue of phosphorylation to be certain that it was residue T208.

Using SEC-MALLS measuring the molecular mass of each protein peak is possible by combining measurements from light scattering and the differential refractive index. Experiments were performed to examine the FHA-$2^{202-310}$ domain in both its phosphorylated and non-phosphorylated states. Initially experiments were performed with $PknF^{1-292}$ although this proved problematic due to the fact that the kinase was still eluting from the column when the FHA-2 domain began to elute. Consequently GST-$PknF^{1-292}$ was expressed and purified from *E. coli* and because this protein was much larger it did not elute from the column at the same time as the FHA-2 domain. Furthermore, after phosphorylation glutathione-s-transferase resin was added to the phosphorylation reaction and incubated with rotation for one hour at 4 ^{o}C in order for the GST-PknF to bind back to the resin. However, if some kinase remained in the reaction then it would probably elute in the void due to the large globular size of the protein. It is suggestive that there was some GST-PknF in the void because the void is much larger in the phosphorylated reaction (figure 6.24a) compared with the non-phosphorylated reaction (figure 6.23a). Furthermore in figure 6.24a a negative refractive index is observed in all protein samples at approximately 40 minutes. The negative refractive index was most probably due to buffer exchange; there was less

salt present in the phosphorylation reactions than in the buffer which conditioned the column.

The results showed that the phosphorylated FHA-2 domain began to elute approximately 2 minutes prior to the non-phosphorylated FHA-2 domain (figure 6.25). The total area under the phosphorylated FHA-2 and non-phosphorylated FHA-2 peaks are approximately the same, thus the concentrations of both proteins are approximately 4 mg/ml; the reason for the smaller peak with the phosphorylated protein is that it eluted over a longer time period and therefore the peak height is smaller. The molecular weight of the phosphorylated FHA-2 domain was shown to be 14.940 kDa at 31.8 minutes and the non-phosphorylated FHA-2 domain had a calculated mass of 10.940 kDa at 33.5 minutes. The molecular mass calculated from the electrospray mass spectroscopy of non-phosphorylated FHA-2 was 11976.5 Da and phosphorylated FHA-2 was 12055.9 Da. The SEC-MALLS results showed that the non-phosphorylated FHA-2 protein was approximately 1 kDa less than the calculated size by mass spectrometry and the phosphorylated FHA-2 was 2.88 kDa larger than the predicted molecular mass of phosphorylated FHA-2. Unfortunately the molecular weights generated are inconclusive and do not offer an explanation as to why the phosphorylated FHA-2 protein eluted before the non-phosphorylated protein (apart from that it is ~ 3 kDa larger). In terms of protein conformation nothing can be drawn from the data. Perhaps only a proportion of the FHA-2 domain was dimeric due to a weak dimerisation and maybe this could account for why the average mass only increased by 2.88 kDa; this is corroborated by the shape of the eluted phosphorylated peak which was not perfectly bell shaped as observed for the non-phosphorylated protein, the phosphorylated peak eluted over a larger area and had a longer tail (figure 6.25).

Unfortunately, the data generated was inconclusive and did not provide an insight into whether the FHA-$2^{202-310}$ domain alters its conformation upon phosphorylation; consequently **no conclusions can safely be drawn regarding the effect that phosphorylation has on the structure of the FHA-2 domain and no hypotheses can be put forward about how this FHA domain may interact with the Rv1747 pT residues to regulate protein function**.

6.9.4 Summary

This chapter has further uncovered the significant roles that the *pknF* and *Rv1747* genes play in *M. tuberculosis* pathogenesis and has demonstrated the critical importance of the FHA-1 domain for Rv1747 protein function. Both the FHA domains were shown to be functionally active in binding both the identified Rv1747 pT epitopes. Furthermore, Rv1747 has been shown to not only be required for growth in the lungs but is also necessary for growth in the spleens. In addition, combining all the data generated in this study suggests that PknF positively regulates the function of the ABC transporter. A mechanism of regulation has been proposed (figure 6.26) although there is still a way to go to fully elucidate the molecular details of how the *pknF-Rv1747* signal transduction system functions in *M. tuberculosis*.

CHAPTER 7. DISCUSSION AND FUTURE WORK

7.1 Discussion

When this study began little was known about the function of the PknF-Rv1747 signalling system; it was recognised that Rv1747 was required for a virulent infection, furthermore both FHA domains and phosphorylation of the PknF activation loop seemed to be required for the PknF-Rv1747 interaction (Curry *et al.*, 2005; Molle *et al.*, 2004). However, the substrate of Rv1747 and the details of how the signalling system functioned in terms of protein phosphorylation and the importance of particular domains were unknown. This study set out to further define the roles of PknF and Rv1747 in relation to the virulence of *M. tuberculosis* with the view to explore whether either of these proteins or particular domains of the proteins might be suitable drug targets for new interventions, or whether either of the mutants might be useful as part of a novel live attenuated vaccine strain.

This study has increased our knowledge of how the PknF-Rv1747 signalling system functions in *M. tuberculosis*. We now know that PknF does indeed play a role in a signal transduction system with Rv1747 and have further uncovered the significant roles that the *pknF* and *Rv1747* genes play in *M. tuberculosis* pathogenesis. Rv1747 codes for an ABC transporter required for growth in mice that is presumed to export a cellular component. This study has determined that this function is positively controlled by PknF. This kinase interacts with the two FHA domains of Rv1747 in a phosphorylation dependent manner and FHA-1 mutation results in attenuation in macrophages demonstrating the critical importance of this domain for Rv1747 protein function. PknF mediated phosphorylation of two threonine residues on Rv1747 was shown to be important for the ABC transporter function in both macrophage and mouse infection models. Furthermore, both FHA domains bound each Rv1747 phosphopeptide with similar affinities providing a description of the nature of the inter- and intra-molecular protein interactions which control the function of the ABC transporter (figure 6.26).

7.2 Future work and implications for novel *M. tuberculosis* treatment interventions

This study is the first to identify two threonine residues on Rv1747 that are phosphorylated, at least *in vitro*, by PknF. Many factors require investigation to further define the roles of PknF and Rv1747. Firstly, identification of the Rv1747 phosphothreonine residues *in vivo* is of upmost importance to confirm that the residues identified in this study are indeed the ones phosphorylated inside the *M. tuberculosis* cell, although data generated in this investigation has already demonstrated the important role that these residues play in controlling the function of the ABC transporter. As the function of Rv1747 is only required for the *in vivo* growth and survival of *M. tuberculosis* this offered a potential explanation to why no substrate for Rv1747 was elucidated in *in vitro* experiments and also why both the mutants demonstrated no differences in susceptibility to any of the drug and stresses investigated.

For PknF, it is unknown whether a dimerisation event occurs to activate the kinase as observed for the solved kinase domain structures of PknB, PknD and PknE (Gay *et al.*, 2006; Greenstein *et al.*, 2007; Mieczkowski *et al.*, 2008; Ortiz-Lombardia *et al.*, 2003; Scherr *et al.*, 2007; Young *et al.*, 2003). The predicted PknF extracellular domain contains regions of low complexity and lacks a ligand binding domain with a recognisable function (Greenstein *et al.*, 2007). It would therefore be of great interest to determine the structure of the PknF kinase domain to enable a further understanding of how this kinase functions in *M. tuberculosis*. It also has to be considered that *in vivo* PknF may be activated by trans-phosphorylation by another STPK and not necessarily activated by sensing a direct signal. However, the presence of an extracellular domain suggests that this protein does perform a sensory role although this study failed to elucidate the nature of the ligand. Also of considerable importance to further understand this signalling system would be to generate a *pknF-Rv1747* double deletion mutant; characterising the phenotypes of this double mutant may aid in understanding how the system functions.

The experiments performed within this study demonstrated the critical importance of the Rv1747 FHA-1 domain for ABC transporter function. Therefore it is also important to

put the FHA domain mutant strains through a mouse aerosol infection model and investigate whether the FHA-1 S47A strain has a growth attenuation phenotype as observed in macrophages. In addition, the *ΔpknF* mutant should also be characterised in the same infection model to determine whether this gene is required for the growth of *M. tuberculosis in vivo*. The prediction is that the *pknF* mutant would have at least some growth attenuation phenotype *in vivo* as this kinase positively regulates the activity of Rv1747.

Results from this study do not corroborate findings by Deol *et al*. who reported that PknF negatively regulates glucose uptake; using antisense RNA to reduce the levels of PknF protein in *M. tuberculosis* they showed increased glucose uptake 16-fold higher compared with wild type and the antisense strain also grew 1.5 fold faster (Deol *et al.*, 2005). This study has shown that that the *ΔpknF* deletion strain grows at the same rate as wild type *in vitro*; also the phenotype microarrays showed the *ΔpknF* and *ΔRv1747* deletion strains utilised glucose as well as the wild type strain. One might have expected that if PknF functioned to negatively regulate glucose uptake then glucose would have been used significantly more in the *ΔpknF* deletion strain compared to wild type; this was not the case. Furthermore, using the *ΔpknF* mutant in radioactive nutrient uptake assays it was not possible to repeat the results of the Deol study (R. Whalan, Personal Communication).

Overall, the results of this study conclude that the PknF-Rv1747 signalling system is essential for the full virulence of *M. tuberculosis* in mice and macrophages. Further understanding of this relationship may allow the identification of potential new antimycobacterial drugs. In terms of novel therapies, the inhibition of essential signalling pathways in *M. tuberculosis* is an attractive prospect to yield novel classes of drug targets, however to date no kinase inhibitor has been discovered which has a high potency against *M. tuberculosis* despite appearing effective in *in vitro* assays (K. Lougheed, unpublished data; Magnet *et al.*, 2010; Wehenkel *et al.*, 2006).

Perhaps a better potential new drug target for *M. tuberculosis* treatment intervention would be to identify small molecule inhibitors against FHA domain-containing proteins. Experiments in this study demonstrated that the FHA-1 domain is critical for Rv1747

function; therefore small molecule inhibitors could be identified against the domain to inhibit it binding a phosphothreonine epitope. As with protein kinase inhibitors, any such compound would have to be target-specific and not inhibit human FHA domain-containing proteins. Although no such studies into the development of FHA domain inhibitors have been published to date, there have been a couple of studies reporting the discovery of specific inhibitors to a Polo-box domain (PBD) protein, which like FHA domains functions to bind phosphothreonine (and phosphoserine) containing epitopes (Reindl *et al.*, 2008; Reindl *et al.*, 2009). The PBD in question forms part of a serine/threonine kinase called Polo-like kinase 1 (Plk1) which is overexpressed in many types of cancerous human tumours and so inhibitors of this protein would be promising in the treatment of cancer patients (Reindl *et al.*, 2008).

Finally, due to the *ΔRv1747* deletion strain being attenuated for growth in the lungs and spleens of a murine model of infection, a mutation of this gene may be a good candidate for a component of a novel live attenuated *M. tuberculosis* vaccine strain.

REFERENCES

Alderwick, L. J., Molle, V., Kremer, L., Cozzone, A. J., Dafforn, T. R., Besra, G. S. & Futterer, K. (2006). Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **103**, 2558-2563.

Alexander, K. A., Laver, P. N., Michel, A. L., Williams, M., van Helden, P. D., Warren, R. M. & Gey van Pittius, N. C. (2010). Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg Infect Dis* **16**, 1296-1299.

Alland, D., Kramnik, I., Weisbrod, T. R., Otsubo, L., Cerny, R., Miller, L. P., Jacobs, W. R., Jr. & Bloom, B. R. (1998). Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): the effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **95**, 13227-13232.

Alland, D., Steyn, A. J., Weisbrod, T., Aldrich, K. & Jacobs, W. R., Jr. (2000). Characterization of the *Mycobacterium tuberculosis iniBAC* promoter, a promoter that responds to cell wall biosynthesis inhibition. *J Bacteriol* **182**, 1802-1811.

Amrein, K. E., Takacs, B., Stieger, M., Molnos, J., Flint, N. A. & Burn, P. (1995). Purification and characterization of recombinant human p50csk protein-tyrosine kinase from an *Escherichia coli* expression system overproducing the bacterial chaperones GroES and GroEL. *Proc Natl Acad Sci U S A* **92**, 1048-1052.

Andersen, P. & Doherty, T. M. (2005). The success and failure of BCG - implications for a novel tuberculosis vaccine. *Nat Rev Microbiol* **3**, 656-662.

Antoine, D., Maguire, H. & Story, A. (2006). Epidemiology and response to the growing problem of tuberculosis in London. *Euro Surveill* **11**, 25-28.

Armstrong, J. A. & Hart, P. D. (1971). Response of cultured macrophages to *Mycobacterium tuberculosis*, with observations on fusion of lysosomes with phagosomes. *J Exp Med* **134**, 713-740.

Arora, G., Sajid, A., Gupta, M., Bhaduri, A., Kumar, P., Basu-Modak, S. & Singh, Y. (2010). Understanding the Role of PknJ in *Mycobacterium tuberculosis*: Biochemical Characterization and Identification of Novel Substrate Pyruvate Kinase A. *PLoS ONE* **5**, e10772.

Austin, P. E., McCulloch, E. A. & Till, J. E. (1971). Characterization of the factor in L-cell conditioned medium capable of stimulating colony formation by mouse marrow cells in culture. *Journal of Cellular Physiology* **77**, 121-133.

Av-Gay, Y. & Everett, M. (2000). The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol* **8**, 238-244.

Av-Gay, Y. & Deretic, V. (2005). Two-Component Systems, Protein Kinases, and Signal Transduction in *Mycobacterium tuberculosis*. In *Tuberculosis and the Tubercle Bacillus*. Edited by S. T. Cole, K. D. Eisenach, D. N. McMurray & W. R. Jacobs, Jr. Washington: ASM Press.

Bach, H., Wong, D. & Av-Gay, Y. (2009). *Mycobacterium tuberculosis* PtkA is a novel protein tyrosine kinase whose substrate is PtpA. *Biochem J* **420**, 155-160.

Bakal, C. J. & Davies, J. E. (2000). No longer an exclusive club: eukaryotic signalling domains in bacteria. *Trends Cell Biol* **10**, 32-38.

Balganesh, M., Kuruppath, S., Marcel, N., Sharma, S., Nair, A. & Sharma, U. (2010). Rv1218c, an ABC transporter of *Mycobacterium tuberculosis* with implications in drug discovery. *Antimicrob Agents Chemother* **54**, 5167-5172.

Behr, M. A. & Small, P. M. (1999). A historical and molecular phylogeny of BCG strains. *Vaccine* **17**, 915-922.

Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. & Small, P. M. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520-1523.

Bhatt, A., Fujiwara, N., Bhatt, K., Gurcha, S. S., Kremer, L., Chen, B., Chan, J., Porcelli, S. A., Kobayashi, K., Besra, G. S. & Jacobs, W. R., Jr. (2007). Deletion of *kasB* in *Mycobacterium tuberculosis* causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. *Proc Natl Acad Sci U S A* **104**, 5157-5162.

Billman-Jacobe, H. (2004). Glycopeptidolipid synthesis in mycobacteria. *Current Science* **86**, 111-114.

Boitel, B., Ortiz-Lombardia, M., Duran, R., Pompeo, F., Cole, S. T., Cervenansky, C. & Alzari, P. M. (2003). PknB kinase activity is regulated by phosphorylation in two Thr residues and dephosphorylation by PstP, the cognate phospho-Ser/Thr phosphatase, in *Mycobacterium tuberculosis*. *Mol Microbiol* **49**, 1493-1508.

Braibant, M., Gilot, P. & Content, J. (2000). The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiology Reviews* **24**, 449-467.

Brennan, M. J. (2005). The tuberculosis vaccine challenge. *Tuberculosis (Edinb)* **85**, 7-12.

Brennan, M. J., Fruth, U., Milstien, J., Tiernan, R., de Andrade Nishioka, S. & Chocarro, L. (2007). Development of new tuberculosis vaccines: a global perspective on regulatory issues. *PLoS Med* **4**, e252.

Brennan, P. J. & Nikaido, H. (1995). The envelope of mycobacteria. *Annu Rev Biochem* **64**, 29-63.

Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L. M., Pym, A. S., Samper, S., van Soolingen, D. & Cole, S. T. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* **99**, 3684-3689.

Brown, L. (1932). Robert Koch. *Bull N Y Acad Med* **8**, 558-584.

Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B. & Guilhot, C. (1999). Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol Microbiol* **34**, 257-267.

Camacho, L. R., Constant, P., Raynaud, C., Laneelle, M. A., Triccas, J. A., Gicquel, B., Daffe, M. & Guilhot, C. (2001). Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium*

tuberculosis. Evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem* **276**, 19845-19854.

Camus, J. C., Pryor, M. J., Medigue, C. & Cole, S. T. (2002). Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**, 2967-2973.

Canova, M. J., Veyron-Churlet, R., Zanella-Cleon, I., Cohen-Gonsaud, M., Cozzone, A. J., Becchi, M., Kremer, L. & Molle, V. (2008). The *Mycobacterium tuberculosis* serine/threonine kinase PknL phosphorylates Rv2175c: mass spectrometric profiling of the activation loop phosphorylation sites and their role in the recruitment of Rv2175c. *Proteomics* **8**, 521-533.

Canova, M. J., Kremer, L. & Molle, V. (2009). The *Mycobacterium tuberculosis* GroEL1 chaperone is a substrate of Ser/Thr Protein Kinases. *J Bacteriol* **191**, 2876-2883.

Chao, J., Wong, D., Zheng, X., Poirier, V., Bach, H., Hmama, Z. & Av-Gay, Y. (2010). Protein kinase and phosphatase signaling in *Mycobacterium tuberculosis* physiology and pathogenesis. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* **In Press, Corrected Proof**.

Colangeli, R., Helb, D., Sridharan, S., Sun, J., Varma-Basil, M., Hazbon, M. H., Harbacheuski, R., Megjugorac, N. J., Jacobs, W. R., Holzenburg, A., Sacchettini, J. C. & Alland, D. (2005). The *Mycobacterium tuberculosis iniA* gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. *Mol Microbiol* **55**, 1829-1840.

Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V. & Mosteller, F. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *Jama* **271**, 698-702.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. & Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544.

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J. R. & Barrell, B. G. (2001). Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007-1011.

Cole, S. T. (2002). Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology* **148**, 2919-2928.

Collins, L. & Franzblau, S. G. (1997). Microplate alamar blue assay versus BACTEC 460 system for high-throughput screening of compounds against *Mycobacterium tuberculosis* and *Mycobacterium avium*. *Antimicrob Agents Chemother* **41**, 1004-1009.

Cowley, S., Ko, M., Pick, N., Chow, R., Downing, K. J., Gordhan, B. G., Betts, J. C., Mizrahi, V., Smith, D. A., Stokes, R. W. & Av-Gay, Y. (2004). The *Mycobacterium tuberculosis* protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth *in vivo*. *Mol Microbiol* **52**, 1691-1702.

Cozzone, A. J. (2005). Role of protein phosphorylation on serine/threonine and tyrosine in the virulence of bacterial pathogens. *J Mol Microbiol Biotechnol* **9**, 198-213.

Curry, J. M., Whalan, R., Hunt, D. M., Gohil, K., Strom, M., Rickman, L., Colston, M. J., Smerdon, S. J. & Buxton, R. S. (2005). An ABC transporter containing a forkhead-associated domain interacts with a serine-threonine protein kinase and is required for growth of *Mycobacterium tuberculosis* in mice. *Infect Immun* **73**, 4471-4477.

Danilchanka, O., Mailaender, C. & Niederweis, M. (2008). Identification of a novel multidrug efflux pump of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **52**, 2503-2511.

Dasgupta, A., Datta, P., Kundu, M. & Basu, J. (2006). The serine/threonine kinase PknB of *Mycobacterium tuberculosis* phosphorylates PBPA, a penicillin-binding protein required for cell division. *Microbiology* **152**, 493-504.

Dasgupta, A., Sureka, K., Mitra, D., Saha, B., Sanyal, S., Das, A. K., Chakrabarti, P., Jackson, M., Gicquel, B., Kundu, M. & Basu, J. (2010). An oligopeptide transporter of *Mycobacterium tuberculosis* regulates cytokine release and apoptosis of infected macrophages. *PLoS ONE* **5**, e12225.

Dassa, E. & Bouige, P. (2001). The ABC of ABCS: a phylogenetic and functional classification of ABC systems in living organisms. *Res Microbiol* **152**, 211-229.

Dassa, E. (2003). Phylogenetic and functional classification of ABC (ATP-binding casssette) systems. In *ABC Proteins from bacteria to man*, pp. 3-35. Edited by I. B. Holland, S. P. C. Cole, K. Kuchler & C. F. Higgins: Elsevier Science.

Davidson, A. L. & Maloney, P. C. (2007). ABC transporters: how small machines do a big job. *Trends in Microbiology* **15**, 448-455.

Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. (2008). Structure, function, and evolution of Bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* **72**, 317-364.

Dawson, R. J., Hollenstein, K. & Locher, K. P. (2007). Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism. *Mol Microbiol* **65**, 250-257.

de Carvalho, L. P., Zhao, H., Dickinson, C. E., Arango, N. M., Lima, C. D., Fischer, S. M., Ouerfelli, O., Nathan, C. & Rhee, K. Y. Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase. *Chem Biol* **17**, 323-332.

Deol, P., Vohra, R., Saini, A. K., Singh, A., Chandra, H., Chopra, P., Das, T. K., Tyagi, A. K. & Singh, Y. (2005). Role of *Mycobacterium tuberculosis* Ser/Thr Kinase PknF: Implications in glucose transport and cell division. *J Bacteriol* **187**, 3415-3420.

Dhandayuthapani, S., Mudd, M. & Deretic, V. (1997). Interactions of OxyR with the promoter region of the oxyR and ahpC genes from *Mycobacterium leprae* and *Mycobacterium tuberculosis*. *J Bacteriol* **179**, 2401-2409.

Domenech, P., Kobayashi, H., LeVier, K., Walker, G. C. & Barry, C. E., 3rd (2009). BacA, an ABC transporter involved in maintenance of chronic murine infections with *Mycobacterium tuberculosis*. *J Bacteriol* **191**, 477-485.

Dover, L. G., Alahari, A., Gratraud, P., Gomes, J. M., Bhowruth, V., Reynolds, R. C., Besra, G. S. & Kremer, L. (2007). EthA, a common activator of thiocarbamide-containing drugs acting on different mycobacterial targets. *Antimicrob Agents Chemother* **51**, 1055-1063.

Duran, R., Villarino, A., Bellinzoni, M., Wehenkel, A., Fernandez, P., Boitel, B., Cole, S. T., Alzari, P. M. & Cervenansky, C. (2005). Conserved autophosphorylation pattern in activation loops and juxtamembrane regions of *Mycobacterium tuberculosis* Ser/Thr protein kinases. *Biochemical and Biophysical Research Communications* **333**, 858-867.

Durocher, D., Taylor, I. A., Sarbassova, D., Haire, L. F., Westcott, S. L., Jackson, S. P., Smerdon, S. J. & Yaffe, M. B. (2000). The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol Cell* **6**, 1169-1182.

Durocher, D. & Jackson, S. P. (2002). The FHA domain. *FEBS Lett* **513**, 58-66.

Durocher, D. (2003). Bacterial signal transduction: a FHAscinating glimpse at the origins of phospho-dependent signal transduction. *Trends Microbiol* **11**, 67-68.

England, P., Wehenkel, A., Martins, S., Hoos, S., Andre-Leroux, G., Villarino, A. & Alzari, P. M. (2009). The FHA-containing protein GarA acts as a phosphorylation-dependent molecular switch in mycobacterial signaling. *FEBS Lett* **583**, 301-307.

Farhana, A., Kumar, S., Rathore, S. S., Ghosh, P. C., Ehtesham, N. Z., Tyagi, A. K. & Hasnain, S. E. (2008). Mechanistic insights into a novel exporter-importer system of *Mycobacterium tuberculosis* unravel its role in trafficking of iron. *PLoS ONE* **3**, e2087.

Fletcher, H. & McShane, H. (2006). Tuberculosis vaccines: current status and future prospects. *Expert Opin Emerg Drugs* **11**, 207-215.

Fletcher, H. A., Hawkridge, T. & McShane, H. (2009). A new vaccine for tuberculosis:the challenges of development and deployment. *J Bioeth Inq* **6**, 219-228.

Gay, L. M., Ng, H. L. & Alber, T. (2006). A conserved dimer and global conformational changes in the structure of apo-PknE Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *J Mol Biol* **360**, 409-420.

Gopalaswamy, R., Narayanan, S., Jacobs, W. R. & Av-Gay, Y. (2007). *Mycobacterium smegmatis* biofilm formation and sliding motility are affected by the serine/threonine protein kinase PknF. *FEMS Microbiol Lett*, 1-7.

Gopalaswamy, R., Narayanan, S., Chen, B., Jacobs, W. R. & Av-Gay, Y. (2009). The serine/threonine protein kinase PknI controls the growth of *Mycobacterium tuberculosis* upon infection. *FEMS Microbiol Lett* **295**, 23-29.

Greenstein, A. E., Grundner, C., Echols, N., Gay, L. M., Lombana, T. N., Miecskowski, C. A., Pullen, K. E., Sung, P. Y. & Alber, T. (2005). Structure/function studies of Ser/Thr and Tyr protein phosphorylation in *Mycobacterium tuberculosis*. *J Mol Microbiol Biotechnol* **9**, 167-181.

Greenstein, A. E., Echols, N., Lombana, T. N., King, D. S. & Alber, T. (2007a). Allosteric Activation by Dimerization of the PknD receptor Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *J Biol Chem* **282**, 11427-11435.

Greenstein, A. E., MacGurn, J. A., Baer, C. E., Falick, A. M., Cox, J. S. & Alber, T. (2007b). *M. tuberculosis* Ser/Thr protein kinase D phosphorylates an anti-anti-sigma factor homolog. *PLoS Pathog* **3**, e49.

Grundner, C., Gay, L. M. & Alber, T. (2005). *Mycobacterium tuberculosis* serine/threonine kinases PknB, PknD, PknE, and PknF phosphorylate multiple FHA domains. *Protein Science* **14**, 1918-1921.

Guilhot, C., Jackson, M. & Gicquel, B. (1999). Mobile genetic elements and plasmids:tools for genetic studies. In *Mycobacteria; Molecular Biology and Virulence*, pp. 17-37. Edited by C. Ratledge & J. Dale: Blackwell Science Ltd.

Gupta, M., Sajid, A., Arora, G., Tandon, V. & Singh, Y. (2009). Forkhead-associated domain-containing protein Rv0019c and polyketide-associated protein PapA5, from substrates of serine/threonine protein kinase PknB to interacting proteins of *Mycobacterium tuberculosis*. *J Biol Chem* **284**, 34723-34734.

Hammet, A., Pike, B., McNees, C., Conlan, L., Tenis, N. & Heierhorstrg, J. (2003). FHA Domains as phospho-threonine binding modules in cell signaling. *IUBMB Life* **55**, 23-27.

Hanks, S. K., Quinn, A. M. & Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42-52.

Hanks, S. K. (1991). Eukaryotic protein kinases. *Curr Opin Struct Biol* **1**, 369-383.

Hanks, S. K. & Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *Faseb J* **9**, 576-596.

Hart, P. D. & Sutherland, I. (1977). BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. *Br Med J* **2**, 293-295.

Hegymegi-Barakonyi, B., Szekely, R., Varga, Z., Kiss, R., Borbely, G., Nemeth, G., Banhegyi, P., Pato, J., Greff, Z., Horvath, Z., Meszaros, G., Marosfalvi, J., Eros, D., Szantai-Kis, C., Breza, N., Garavaglia, S., Perozzi, S., Rizzi, M., Hafenbradl, D., Ko, M., Av-Gay, Y., Klebl, B. M., Orfi, L. & Keri, G. (2008). Signalling inhibitors against *Mycobacterium tuberculosis*: early days of a new therapeutic concept in tuberculosis. *Curr Med Chem* **15**, 2760-2770.

Higgins, C. F., Hiles, I. D., Salmond, G. P., Gill, D. R., Downie, J. A., Evans, I. J., Holland, I. B., Gray, L., Buckel, S. D., Bell, A. W. & et al. (1986). A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature* **323**, 448-450.

Higgins, C. F. (2001). ABC transporters: physiology, structure and mechanism--an overview. *Res Microbiol* **152**, 205-210.

Higgins, C. F. & Linton, K. J. (2003). ABC transporters: An introduction and overview. In *ABC proteins: from bacteria to man*, pp. xxvii-xxiii. Edited by I. B. Holland, S. P. C. Cole, K. Kuchler & C. F. Higgins: Academic Press.

Hinds, J., Mahenthiralingam, E., Kempsell, K. E., Duncan, K., Stokes, R. W., Parish, T. & Stoker, N. G. (1999). Enhanced gene replacement in mycobacteria. *Microbiology* **145**, 519-527.

Hofmann, K. & Bucher, P. (1995). The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem Sci* **20**, 347-349.

Homerova, D., Halgasova, L. & Kormanec, J. (2008). Cascade of extracytoplasmic function sigma factors in *Mycobacterium tuberculosis*: identification of a sigmaJ-dependent promoter upstream of *sigI*. *FEMS Microbiol Lett* **280**, 120-126.

Horwich, A. L., Farr, G. W. & Fenton, W. A. (2006). GroEL-GroES-mediated protein folding. *Chem Rev* **106**, 1917-1930.

Jacobs, W. R., Kalpana, G. V., Cirillo, J. D., Pascopella, L., Snapper, S. B., Udani, R. A., Jones, W., Barletta, R. G., Bloom, B. R. & Jeffrey, H. M. (1991). Genetic systems for mycobacteria. In *Methods Enzymol*, pp. 537-555: Academic Press.

Jackson, M., Berthet, F. X., Otal, I., Rauzier, J., Martin, C., Gicquel, B. & Guilhot, C. (1996). The *Mycobacterium tuberculosis* purine biosynthetic pathway: isolation and characterization of the *purC* and *purL* genes. *Microbiology* **142**, 2439-2447.

Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. J. (2005). The adaptive immune response. In *Immunobiology; The Immune System in Health and Disease*. Edited by C. A. Janeway, P. Travers, M. Walport & M. J. Shlomchik.

Jang, J., Stella, A., Boudou, F., Levillain, F., Darthuy, E., Vaubourgeix, J., Wang, C., Bardou, F., Puzo, G., Gilleron, M., Burlet-Schiltz, O., Monsarrat, B., Brodin, P., Gicquel, B. & Neyrolles, O. (2010). Functional characterization of the *Mycobacterium tuberculosis* serine/threonine kinase PknJ. *Microbiology* **156**, 1619-1631.

Jayakumar, D., Jacobs, W. R., Jr. & Narayanan, S. (2008). Protein kinase E of *Mycobacterium tuberculosis* has a role in the nitric oxide stress response and apoptosis in a human macrophage model of infection. *Cell Microbiol* **10**, 365-374.

Kang, C. M., Abbott, D. W., Park, S. T., Dascher, C. C., Cantley, L. C. & Husson, R. N. (2005). The *Mycobacterium tuberculosis* serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. *Genes Dev* **19**, 1692-1704.

Khan, S., Nagarajan, S. N., Parikh, A., Samantaray, S., Singh, A., Kumar, D., Roy, R. P., Bhatt, A. & Nandicoori, V. K. (2010). Phosphorylation of enoyl-ACP reductase InhA impacts mycobacterial growth and survival. *J Biol Chem* **285**, 37860-37871.

Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc Natl Acad Sci U S A* **103**, 17783-17788.

Koul, A., Choidas, A., Treder, M., Tyagi, A. K., Drlica, K., Singh, Y. & Ullrich, A. (2000). Cloning and characterization of secretory tyrosine phosphatases of *Mycobacterium tuberculosis*. *J Bacteriol* **182**, 5425-5432.

Koul, A., Choidas, A., Tyagi, A. K., Drlica, K., Singh, Y. & Ullrich, A. (2001). Serine/threonine protein kinases PknF and PknG of *Mycobacterium tuberculosis*: characterization and localization. *Microbiology* **147**, 2307-2314.

Kremer, L. & Besra, G. S. (2005). A Waxy Tale, by *Mycobacterium tuberculosis*. In *Tuberculosis and the Tubercle Bacillus*, pp. 287-305. Edited by S. P. C. Cole, K. D. Eisenach, D. N. McMurray & W. R. Jacobs, Jr.: ASM Press, Washington, D. C.

Lakshminarayan, H., Narayanan, S., Bach, H., Sundaram, K. G. P. & Av-Gay, Y. (2008). Molecular cloning and biochemical characterization of a serine threonine protein kinase, PknL, from *Mycobacterium tuberculosis*. *Prot Express Purificat* **58**, 309-317.

Lalloo, U. G. & Ambaram, A. (2010). New antituberculous drugs in development. *Curr HIV/AIDS Rep* **7**, 143-151.

Lapenna, S. & Giordano, A. (2009). Cell cycle kinases as therapeutic targets for cancer. *Nat Rev Drug Discov* **8**, 547-566.

Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J. & Bork, P. (2006). SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* **34**, D257-260.

Letunic, I., Doerks, T. & Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res* **37**, D229-D232.

Li, J., Taylor, I. A., Lloyd, J., Clapperton, J. A., Howell, S., MacMillan, D. & Smerdon, S. J. (2008). Chk2 oligomerization studied by phosphopeptide ligation. *J Biol Chem* **283**, 36019-36030.

Liang, X. & Van Doren, S. R. (2008). Mechanistic insights into phosphoprotein-binding FHA domains. *Acc Chem Res* **41**, 991-999.

Linton, K. J. & Higgins, C. F. (1998). The *Escherichia coli* ATP-binding cassette (ABC) proteins. *Mol Microbiol* **28**, 5-13.

Lochhead, P. A. (2009). Protein kinase activation loop autophosphorylation in cis: overcoming a Catch-22 situation. *Sci Signal* **2**, pe4.

Madec, E., Stensballe, A., Kjellstrom, S., Cladiere, L., Obuchowski, M., Jensen, O. N. & Seror, S. J. (2003). Mass spectrometry and site-directed mutagenesis identify several autophosphorylated residues required for the activity of PrkC, a Ser/Thr kinase from *Bacillus subtilis*. *J Mol Biol* **330**, 459-472.

Madigan, M. T., Martinko, J., M & Parker, J. (2003). *Brock Biology of Microorganisms*: Prentice Hall.

Magnet, S., Hartkoorn, R. C., Székely, R., Pató, J., Triccas, J. A., Schneider, P., Szántai-Kis, C., Orfi, L., Chambon, M., Banfi, D., Bueno, M., Turcatti, G., Kéri, G. & Cole, S. T. (2010). Leads for antitubercular compounds from kinase inhibitor library screens. *Tuberculosis* **90**, 354-360.

Mahajan, A., Yuan, C., Lee, H., Chen, E. S., Wu, P. Y. & Tsai, M. D. (2008). Structure and function of the phosphothreonine-specific FHA domain. *Sci Signal* **1**, re12.

Malhotra, V., Arteaga-Cortes, L. T., Clay, G. & Clark-Curtiss, J. E. (2010). *Mycobacterium tuberculosis* Protein Kinase K confers survival advantage during early infection in mice and regulates growth in culture and during persistent infection: implications in immune modulation. *Microbiology* **156**, 2829-2841.

Manning, G., Plowman, G. D., Hunter, T. & Sudarsanam, S. (2002a). Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* **27**, 514-520.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. (2002b). The protein kinase complement of the human genome. *Science* **298**, 1912-1934.

Marri, P. R., Bannantine, J. P. & Golding, G. B. (2006). Comparative genomics of metabolic pathways in *Mycobacterium* species: gene duplication, gene decay and lateral gene transfer. *FEMS Microbiol Rev* **30**, 906-925.

Master, S. S., Springer, B., Sander, P., Boettger, E. C., Deretic, V. & Timmins, G. S. (2002). Oxidative stress response genes in *Mycobacterium tuberculosis*: role of *ahpC* in resistance to peroxynitrite and stage-specific survival in macrophages. *Microbiology* **148**, 3139-3144.

McShane, H., Pathan, A. A., Sander, C. R., Keating, S. M., Gilbert, S. C., Huygen, K., Fletcher, H. A. & Hill, A. V. (2004). Recombinant modified vaccinia virus Ankara expressing antigen 85A boosts BCG-primed and naturally acquired antimycobacterial immunity in humans. *Nat Med* **10**, 1240-1244.

McShane, H., Pathan, A. A., Sander, C. R., Goonetilleke, N. P., Fletcher, H. A. & Hill, A. V. (2005). Boosting BCG with MVA85A: the first candidate subunit vaccine for tuberculosis in clinical trials. *Tuberculosis (Edinb)* **85**, 47-52.

Mieczkowski, C., Iavarone, A. T. & Alber, T. (2008). Auto-activation mechanism of the *Mycobacterium tuberculosis* PknB receptor Ser/Thr kinase. *Embo J* **27**, 3186-3197.

Miller, J. H. (1972). *Experiments in Molecular Genetics*: Cold Spring Harbor Press, New York.

Mizuno, T. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res* **4**, 161-168.

Molle, V., Kremer, L., Girard-Blanc, C., Besra, G. S., Cozzone, A. J. & Prost, J. F. (2003). An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *Biochemistry* **42**, 15300-15309.

Molle, V., Soulat, D., Jault, J. M., Grangeasse, C., Cozzone, A. J. & Prost, J. F. (2004). Two FHA domains on an ABC transporter, Rv1747, mediate its phosphorylation by PknF, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *FEMS Microbiol Lett* **234**, 215-223.

Molle, V., Brown, A. K., Besra, G. S., Cozzone, A. J. & Kremer, L. (2006a). The condensing activities of the *Mycobacterium tuberculosis* type II fatty acid synthase are differentially regulated by phosphorylation. *J Biol Chem* **281**, 30094-30103.

Molle, V., Zanella-Cleon, I., Robin, J. P., Mallejac, S., Cozzone, A. J. & Becchi, M. (2006b). Characterization of the phosphorylation sites of *Mycobacterium tuberculosis* serine/threonine protein kinases, PknA, PknD, PknE, and PknH by mass spectrometry. *Proteomics* **6**, 3754-3766.

Narayan, A., Sachdeva, P., Sharma, K., Saini, A. K., Tyagi, A. K. & Singh, Y. (2007). Serine threonine protein kinases of mycobacterial genus: phylogeny to function. *Physiol Genomics* **29**, 66-75.

Niebisch, A., Kabus, A., Schultz, C., Weil, B. & Bott, M. (2006). Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the OdhI protein. *J Biol Chem* **281**, 12300-12307.

Nolen, B., Taylor, S. & Ghosh, G. (2004). Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell* **15**, 661-675.

Nott, T. J., Kelly, G., Stach, L., Li, J., Westcott, S., Patel, D., Hunt, D. M., Howell, S., Buxton, R. S., O'Hare, H. M. & Smerdon, S. J. (2009). An intramolecular switch regulates phosphoindependent FHA domain interactions in *Mycobacterium tuberculosis*. *Science Signaling* **2**, ra12.

O'Hare, H. M., Duran, R., Cervenansky, C., Bellinzoni, M., Wehenkel, A. M., Pritsch, O., Obal, G., Baumgartner, J., Vialaret, J., Johnsson, K. & Alzari, P. M. (2008). Regulation of glutamate metabolism by protein kinases in mycobacteria. *Mol Microbiol* **70**, 1408-1423.

Ortiz-Lombardia, M., Pompeo, F., Boitel, B. & Alzari, P. M. (2003). Crystal Structure of the Catalytic Domain of the PknB serine/threonine kinase from *Mycobacterium tuberculosis*. *J Biol Chem* **278**, 13094-13100.

Pallen, M., Chaudhuri, R. & Khan, A. (2002). Bacterial FHA domains: neglected players in the phospho-threonine signalling game? *Trends Microbiol* **10**, 556-563.

Papavinasasundaram, K. G., Anderson, C., Brooks, P. C., Thomas, N. A., Movahedzadeh, F., Jenner, P. J., Colston, M. J. & Davis, E. O. (2001). Slow induction of RecA by DNA damage in *Mycobacterium tuberculosis*. *Microbiology* **147**, 3271-3279.

Papavinasasundaram, K. G., Chan, B., Chung, J.-H., Colston, M. J., Davis, E. O. & Av-Gay, Y. (2005). Deletion of the *Mycobacterium tuberculosis pknH* gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. *J Bacteriol* **187**, 5751-5760.

Park, S. T., Kang, C. M. & Husson, R. N. (2008). Regulation of the SigH stress response regulon by an essential protein kinase in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **105**, 13105-13110.

Perez, J., Garcia, R., Bach, H., de Waard, J. H., Jacobs, W. R., Jr., Av-Gay, Y., Bubis, J. & Takiff, H. E. (2006). *Mycobacterium tuberculosis* transporter MmpL7 is a potential substrate for kinase PknD. *Biochem Biophys Res Commun* **348**, 6-12.

Pethe, K., Alonso, S., Biet, F., Delogu, G., Brennan, M. J., Locht, C. & Menozzi, F. D. (2001). The heparin-binding haemagglutinin of *M. tuberculosis* is required for extrapulmonary dissemination. *Nature* **412**, 190-194.

Pitarque, S., Larrouy-Maumus, G., Payre, B., Jackson, M., Puzo, G. & Nigou, J. (2008). The immunomodulatory lipoglycans, lipoarabinomannan and lipomannan, are exposed at the mycobacterial cell surface. *Tuberculosis* **88**, 560-565.

Price, C. T. D., Bukka, A., Cynamon, M. & Graham, J. E. (2008). Glycine betaine uptake by the ProXVWZ ABC transporter contributes to the ability of *Mycobacterium tuberculosis* to initiate growth in human macrophages. *J Bacteriol* **190**, 3955-3961.

Prisic, S., Dankwa, S., Schwartz, D., Chou, M. F., Locasale, J. W., Kang, C. M., Bemis, G., Church, G. M., Steen, H. & Husson, R. N. (2010). Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. *Proc Natl Acad Sci U S A* **107**, 7521-7526.

Reindl, W., Yuan, J., Kramer, A., Strebhardt, K. & Berg, T. (2008). Inhibition of polo-like kinase 1 by blocking polo-box domain-dependent protein-protein interactions. *Chem Biol* **15**, 459-466.

Reindl, W., Yuan, J., Kramer, A., Strebhardt, K. & Berg, T. (2009). A pan-specific inhibitor of the polo-box domains of polo-like kinases arrests cancer cells in mitosis. *Chembiochem* **10**, 1145-1148.

Ren, R. (2005). Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer* **5**, 172-183.

Ruiz, N., Kahne, D. & Silhavy, T. J. (2009). Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nat Rev Microbiol* **7**, 677-683.

Russell, D. G. (2001). *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nat Rev Mol Cell Biol* **2**, 569-577.

Russell, D. G. (2007). Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* **5**, 39-47.

Russell, D. G., Barry, C. E., 3rd & Flynn, J. L. (2010). Tuberculosis: what we don't know can, and does, hurt us. *Science* **328**, 852-856.

Salazar, L., Fsihi, H., de Rossi, E., Riccardi, G., Rios, C., Cole, S. T. & Takiff, H. E. (1996). Organization of the origins of replication of the chromosomes of *Mycobacterium smegmatis*, *Mycobacterium leprae* and *Mycobacterium tuberculosis* and isolation of a functional origin from *M. smegmatis*. *Mol Microbiol* **20**, 283-293.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual,*, Third edn: Cold Spring Harbor, New York.

Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77-84.

Saurin, W., Hofnung, M. & Dassa, E. (1999). Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evoln* **48**, 22-41.

Scherr, N., Honnappa, S., Kunz, G., Mueller, P., Jayachandran, R., Winkler, F., Pieters, J. & Steinmetz, M. O. (2007). Structural basis for the specific inhibition of protein kinase G, a virulence factor of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **104**, 12151-12156.

Scherr, N., Muller, P., Perisa, D., Combaluzier, B., Jeno, P. & Pieters, J. (2009). Survival of Pathogenic Mycobacteria in Macrophages Is Mediated through Autophosphorylation of Protein Kinase G. *J Bacteriol* **191**, 4546-4554.

Schneider, E. & Hunke, S. (1998). ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol Rev* **22**, 1-20.

Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *PNAS* **95**, 5857-5864.

Sharma, K., Gupta, M., Krupa, A., Srinivasan, N. & Singh, Y. (2006a). EmbR, a regulatory protein with ATPase activity, is a substrate of multiple serine/threonine kinases and phosphatase in *Mycobacterium tuberculosis*. *Febs J* **273**, 2711-2721.

Sharma, K., Gupta, M., Pathak, M., Gupta, N., Koul, A., Sarangi, S., Baweja, R. & Singh, Y. (2006b). Transcriptional control of the mycobacterial *embCAB* operon by PknH through a regulatory protein, EmbR, *in vivo*. *J Bacteriol* **188**, 2936-2944.

Sherman, D. R., Voskuil, M., Schnappinger, D., Liao, R., Harrell, M. I. & Schoolnik, G. K. (2001). Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha - crystallin. *Proc Natl Acad Sci U S A* **98**, 7534-7539.

Smith, I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev* **16**, 463-496.

Smollett, K. L., Fivian-Hughes, A. S., Smith, J. E., Chang, A., Rao, T. & Davis, E. O. (2009). Experimental determination of translational start sites resolves uncertainties in genomic open reading frame predictions - application to *Mycobacterium tuberculosis*. *Microbiology* **155**, 186-197.

Springer, B., Sander, P., Sedlacek, L., Ellrott, K. & Boettger, E. C. (2001). Instability and site-specific excision of integration-proficient mycobacteriophage L5 plasmids: development of stably maintained integrative vectors. *Int J Med Microbiol* **290**, 669-675.

Steenken, W., Oatway, W. H. & Petroff, S. A. (1934). Biological studies of the tubercle bacillus : Dissociation and pathogenicity of the R and S variants of the human tubercle bacillus (H(37)). *J Exp Med* **60**, 515-540.

Steenken, W. J. & Gardner, L. U. (1946). History of H37Rv strain of tubercle bacillus. *American Review of Tuberculosis* **54**, 62-66.

Stewart, G. R., Robertson, B. D. & Young, D. B. (2003). Tuberculosis: a problem with persistence. *Nat Rev Microbiol* **1**, 97-105.

Strohl, W. R. (1992). Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Res* **20**, 961-974.

Sureka, K., Hossain, T., Mukherjee, P., Chatterjee, P., Datta, P., Kundu, M. & Basu, J. (2010). Novel role of phosphorylation-dependent interaction between FtsZ and FipA in mycobacterial cell division. *PLoS ONE* **5**, e8590.

Szekely, R., Waczek, F., Szabadkai, I., Nemeth, G., Hegymegi-Barakonyi, B., Eros, D., Szokol, B., Pato, J., Hafenbradl, D., Satchell, J., Saint-Joanis, B., Cole, S. T., Orfi, L., Klebl, B. M. & Keri, G. (2008). A novel drug discovery concept for tuberculosis: inhibition of bacterial and host cell signalling. *Immunol Lett* **116**, 225-231.

Teo, S. S. & Shingadia, D. V. (2006). Does BCG have a role in tuberculosis control and prevention in the United Kingdom? *Arch Dis Child* **91**, 529-531.

Thakur, M. & Chakraborti, P. K. (2006). GTPase activity of mycobacterial FtsZ is impaired due to its transphosphorylation by the eukaryotic-type Ser/Thr kinase, PknA. *J Biol Chem* **281**, 40107-40113.

Timm, J., Gomez, M. A. & Smith, I. (1999). Gene expression and regulation. In *Mycobacteria; Molecular Biology and Virulence*, pp. 59-92. Edited by C. Ratledge & J. Dale: Blackwell Science Ltd.

Torrelles, J. B., Khoo, K. H., Sieling, P. A., Modlin, R. L., Zhang, N., Marques, A. M., Treumann, A., Rithner, C. D., Brennan, P. J. & Chatterjee, D. (2004). Truncated structural variants of lipoarabinomannan in *Mycobacterium leprae* and an ethambutol-resistant strain of *Mycobacterium tuberculosis*. *J Biol Chem* **279**, 41227-41239.

Ubersax, J. A. & Ferrell, J. E., Jr. (2007). Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* **8**, 530-541.

Vandal, O. H., Roberts, J. A., Odaira, T., Schnappinger, D., Nathan, C. F. & Ehrt, S. (2009). Acid-susceptible mutants of *Mycobacterium tuberculosis* share hypersusceptibility to cell wall and oxidative stress and to the host environment. *J Bacteriol* **191**, 625-631.

Vanzembergh, F., Peirs, P., Lefevre, P., Celio, N., Mathys, V., Content, J. & Kalai, M. 2010. Effect of PstS sub-units or PknD deficiency on the survival of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **90**, 338-345.

Velamakanni, S., Wei, S. L., Janvilisri, T. & van Veen, H. W. (2007). ABCG transporters: structure, substrate specificities and physiological roles : a brief overview. *J Bioenerg Biomembr* **39**, 465-471.

Veyron-Churlet, R., Molle, V., Taylor, R. C., Brown, A. K., Besra, G. S., Zanella-Cleon, I., Futterer, K. & Kremer, L. (2009). The *Mycobacterium tuberculosis* ß-ketoacyl-acyl carrier protein synthase III activity is inhibited by phosphorylation on a single threonine residue. *J Biol Chem* **284**, 6414-6424.

Villarino, A., Duran, R., Wehenkel, A., Fernandez, P., England, P., Brodin, P., Cole, S. T., Zimny-Arndt, U., Jungblut, P. R., Cervenansky, C. & Alzari, P. M. (2005). Proteomic identification of *M. tuberculosis* protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions. *J Mol Biol* **350**, 953-963.

von Eiff, C., McNamara, P., Becker, K., Bates, D., Lei, X. H., Ziman, M., Bochner, B. R., Peters, G. & Proctor, R. A. (2006). Phenotype microarray profiling of *Staphylococcus aureus menD* and *hemB* mutants with the small-colony-variant phenotype. *J Bacteriol* **188**, 687-693.

Voskuil, M. I., Schnappinger, D., Rutherford, R., Liu, Y. & Schoolnik, G. K. (2004). Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis (Edinb)* **84**, 256-262.

Walburger, A., Koul, A., Ferrari, G., Nguyen, L., Prescianotto-Baschong, C., Huygen, K., Klebl, B., Thompson, C., Bacher, G. & Pieters, J. (2004). Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* **304**, 1800-1804.

Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J* **1**, 945-951.

Walker, K. B., Brennan, M. J., Ho, M. M., Eskola, J., Thiry, G., Sadoff, J., Dobbelaer, R., Grode, L., Liu, M. A., Fruth, U. & Lambert, P. H. (2010). The second Geneva consensus: recommendations for novel live TB vaccines. *Vaccine* **28**, 2259-2270.

Wang, Z. X. & Wu, J. W. (2002). Autophosphorylation kinetics of protein kinases. *Biochem J* **368**, 947-952.

Webb, V. & Davies, J. (1999). Antibiotics and antibiotic resistance in mycobacteria. In *Mycobacteria: Molecular Biology and Virulence*. Edited by C. Ratledge & J. Dale.

Wehenkel, A., Fernandez, P., Bellinzoni, M., Catherinot, V., Barilone, N., Labesse, G., Jackson, M. & Alzari, P. M. (2006). The structure of PknB in complex with mitoxantrone, an ATP-competitive inhibitor, suggests a mode of protein kinase regulation in mycobacteria. *FEBS Lett* **580**, 3018-3022.

Wehenkel, A., Bellinzoni, M., Grana, M., Duran, R., Villarino, A., Fernandez, P., Andre-Leroux, G., England, P., Takiff, H., Cervenansky, C., Cole, S. T. & Alzari, P. M. (2008). Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential. *Biochim Biophys Acta* **1784**, 193-202.

West, A. H. & Stock, A. M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**, 369-376.

WHO (2010). Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 Global Report on Surveillance and Response.

Young, D. & Dye, C. (2006). The development and impact of tuberculosis vaccines. *Cell* **124**, 683-687.

Young, T. A., Delagoutte, B., Endrizzi, J. A., Falick, A. M. & Alber, T. (2003). Structure of *Mycobacterium tuberculosis* PknB supports a universal activation mechanism for Ser/Thr protein kinases. *Nat Struct Biol* **10**, 168-174.

APPENDIX I. MEDIA AND BUFFER COMPOSITION

L-Broth/Agar
10 g Tryptone
5 g Yeast Extract
10 g NaCl
15 g Agar (if required; Difco, Becton Dickinson)
dH_2O to 1 litre
Autoclave at 121 ºC for 15 mins

Modified Dubos Medium
1 g KH_2PO_4
6.25 g $Na_2HPO_4.12H_2O$
1.25 g Na_3 Citrate
0.6 g $MgSO_4.7H_2O$
2 g Asparagine
2 g Casamino Acids
5 ml 10% Tween 80
2 ml Glycerol
dH_2O to 960 ml
Adjust to pH 7.2
Autoclave at 121 ºC for 15 minutes
40 ml Dubos medium Albumin (Difco, Becton Dickinson)

Mycobacteria 7H11 Agar
21 g Mycobacteria 7H11 Agar Powder (Difco, Becton Dickinson)
5 ml Glycerol
dH_2O to 960 ml
Autoclave at 121 ºC for 15 minutes
Add 40 ml or 100 ml Dubos Medium Albumin (Difco, Becton Dickinson)

PBS
10 g NaCl
0.25 g KCl
1.437 g Na_2HPO_4
0.25 g KH_2PO_4
dH_2O to 1 litre
Autoclave at 121 ºC for 15 minutes

Resuspension Buffer
25 mM Tris-HCl pH7.9
10 mM EDTA
50 mM Glucose

Sauton's Medium

0.5 g KH_2PO_4
0.5 g $MgSO_4$
4.0 g L-asparagine
60 ml glycerol
0.05 g Ferric ammonium sulfate
2.0 g Citric acid
0.1 ml 1 % $ZnSO4.7H_2O$
dH_2O to 1 litre
Adjust to pH 7.0
Filter sterilize (0.22 μM)

SSC Buffer (20x)

175.2 g NaCl
88.2 g Tris-Sodium Citrate
dH_2O to 1 litre
Adjust to pH 7.0-7.2
Autoclave at 121 ºC for 15 minutes

TBE Buffer (10X)

121.1 g Tris Base
61.83 g Boric acid
18.6 g EDTA
dH_2O to 1 litre
Adjust to pH 8.0

TE Buffer

1.211 g Tris Base
0.372 g EDTA
dH_2O to 1 litre
Adjust to pH 8.3
Autoclave at 121 ºC for 15 minutes

Z-Buffer (Miller, 1972)

60 mM Na_2HPO_4
40 mM NaH_2PO_4
10 mM KCl
1 mM $MgSO_4$
Adjust to pH 7.0
Autoclave at 121 ºC for 15 minutes
When required add β-mercaptoethanol to a final concentration of 50 mM

APPENDIX II. OLIGONUCLEOTIDES USED IN THIS STUDY

RT-PCR Primer Pairs

Primer Name	Description	Sequence (5'-3')
pknF F	gene specific internal primer	AACATCCTGATCGCCAATCC
pknF R	gene specific internal primer	TTGCAGCGTCGAGCAGTAGG
pknF-Rv1747 F	co-transcription primer	AAGGCACCAACACCACCATCT
pknF-Rv1747 R	co-transcription primer	GGACAGGTTGGGCAGGCGTAT
Rv1747 F	gene specific internal primer	CGTTCACGCCGAATATGCCT
Rv1747 R	gene specific internal primer	CCATGAAGACCGCACCGACA
Rv1747-Rv1748 F	co-transcription primer	AGGATTCGCATTGGCATCAC
Rv1747-Rv1748 R	co-transcription primer	GGCTTGTAGCTTGGCCTTGT
Rv1748 F	gene specific internal primer	GGCGATCTTGCGTCGGATAG
Rv1748 R	gene specific internal primer	GTACGGTCCGGCAACACGAT

qRT-PCR Primer Pairs

Primer Name	Sequence (5'-3')
pknF F	CACGAACGTCGGCTGTTG
pknF R	GACGATCAGGTGAATCAGGATTG
Rv1747 F	TACGGTCGACCTGATCAAATTG
Rv1747 R	GCGCTGGCGGTGTGA
iniA F	TCATCGCAGTCTCATCACTGTTG
iniA R	TTGGACTCTTCGTTGAGCTCTTT
iniB F	TTATCGATTACATCCTGAGCCTGTT
iniB R	CGGAGCGGCAACGAA
iniC F	ACTCCGAATGCTAAGCCTTTTG
iniC R	CAGCGACGCGATTTCGT
ethA F	GCAAGCCCATCCTCGAGTAC
ethA R	CGGATATGCCTGTCGATTCC
pknD F	CAACGGACAGTTCTTTGTCGAA
pknD R	TGTTTCAATAGGGCGCGTAAA
sigA F	TCGGTTCGCGCCTACCT
sigA R	GGCTAGCTCGACCTCTTCCT

Primers for *pknF* 5'RACE

Primer Name	Sequence (5'-3')
5' RACE GSP1	ACCGCAGTGATGATCTC
5' RACE GSP2	GGACGGCGACGATGTGTGGATGC
5' RACE GSP3	CCTGGCGGGGCAGTCTGGGAT

Primer pair for cloning *pknF* promoter region

Primer Name	Restriction site	Sequence (5'-3')
pknF promoter F	*XbaI*	CCTCTAGACTGATCGTGCAGATCGAACA
pknF promoter R	*HindIII*	CCAAGCTTGGCGGCATCGTAGCCCGCCA

Protein expression primers for cloning into pGEX-6P-1

Primer Name	Restriction site	Sequence (5'-3')
PknF $^{1-292}$ F	*BamHI*	GGATCCATGCCGCTCGCGGAAGGTTCG
PknF $^{1-292}$ R	*XhoI* plus STOP	CTCGAGTCACGGTTGCGACACCCGCGT
Rv1747 $^{1-559}$ F	*BamHI*	GGATCCGTGCCGATGAGCCAACCAGCC
Rv1747 $^{1-559}$ R	*EcoRI* plus STOP	GAATTCTCAGTCGTCGGCGACGGTGCTGAA
FHA 1 $^{1-120}$ F	*BamHI plus ATG*	CCGGATCCGTGCCGATGAGCCAACCA
FHA 1 $^{1-120}$ R	*EcoRI* plus STOP	CCGGATTCTCAGCGTATCGACGTCGTCTG
FHA 2 $^{202-310}$ F	*BamHI plus ATG*	CCGGATCCATGACTGAGGCGGGAAACCTC
FHA 2 $^{202-310}$ R	*EcoRI* plus STOP	CCGGATTCTCAGTTCTCTTCACGGCGCGC

Site directed mutagenesis of pVS_03 and pRW76

Primer Name	Sequence (5'-3') (highlighted bases indicate mutation)
Rv1747 T150A F	TACAACAGCTTCCACCGGCCGCCACCCGGATACCCGCCGCTCC
Rv1747 T150A R	GGAGCGGCGGGTATCCGGGTGGCGGCCGGTGGAAGCTGTTGTA
Rv1747 T150D F	TACAACAGCTTCCACCGGCCGACACCCGGATACCCGCCGCTCC
Rv1747 T150D R	GGAGCGGCGGGTATCCGGGTGTCGGCCGGTGGAAGCTGTTGTA
Rv1747 T208A F	CTGAGGCGGGAAACCTCGCGGCATCGATGATGAAGATCCTGCG
Rv1747 T208A R	CGCAGGATCTTCATCATCGATGCCGCGAGGTTTCCCGCCTCAG
Rv1747 T208D F	CTGAGGCGGGAAACCTCGCGGACTCGATGATGAAGATCCTGCG
Rv1747T208D R	CGCAGGATCTTCATCATCGAGTCCGCGAGGTTTCCCGCCTCAG

Primers used for sequencing

Sequencing primers for *E. coli* expression clones		
Primer Name	**Description**	**Sequence (5'-3')**
M13 F	To sequence DNA inserts in pCR4Blunt-TOPO	TGTAAAACGACGGCCAGT
M13 R	To sequence DNA inserts in pCR4Blunt-TOPO	CAGGAAACAGCTATGACC
PGEX 3'	To sequence DNA inserts in pGex6P1	CCGGGAGCTGCATGTGTCAGAGG
PGEX 5'	To sequence DNA inserts in pGex6P1	GGGCTGGCAAGCCACGTTTGGTG
pEJ414 pmint2	To sequence DNA inserts in pEJ414	ACGAGGGGCATTCACACCAGATTG
pEJ414 lacR	To sequence DNA inserts in pEJ414	TTCCCAGTCACGACGTTGTAAAA
pMv306-4		CCAGCAACGCGGCCTTTTTACG
pKP186 F	To sequence DNA inserts in pKP186	
pKP186 R	To sequence DNA inserts in pKP186	
Sequencing primers for site directed mutagenesis conformation		
Rv1747 T150 F	Forward primer to screen for T150 mutation	AACCTGTCCGCGGGAGCGTGGCCC
Rv1747 T150 R	Reverse primer to screen for T150 mutation	GTGGCCACAACTGTTGCCCACC
Rv1747 T208 F	Forward primer to screen for T208 mutation	CCCGCGGCGGAACTGAGGCG
Rv1747 T208 R	Reverse primer to screen for T208 mutation	ACTCCCCCGTCAACCTGCCTG

APPENDIX III. MICROARRAY DATA ANALYSIS

To allow for statistical analysis three biological replicates were performed for each condition. Each biological replicate was performed in technical duplicate with dye-swaps to correct for dye-labelling bias.

Microarray data analysis was performed using the GeneSpring GX 10 software (Agilent Technologies) as follows:

- An advanced workflow analysis was chosen using the generic two colour system
- Data was then loaded into the experiment and the following operations were performed:
    a. The technology (*M. tuberculosis* genome annotation) was chosen: TB_v_2_2_flags_GX10
    b. Excel spreadsheets from Bluefuse output files were loaded
    c. Dye swaps were highlighted. Experiments were ticked in which the samples were labelled in the opposite direction. GeneSpring then reverses these colours and this is how dye bias is corrected for.
    d. Due to dye swaps the LOWESS normalization step was not required. LOWESS normalization is a mathematical correction for dye labelling bias but as this was corrected for using dye-swaps then this correction step was not necessary.
- Parameters were set so that each microarray slide could be identified by biological replicate, technical replicate and strain/genotype.
- Quality control on samples was then performed using principal components analysis. Microarray slides were removed from the analysis at this stage if they were outliers.
- Quality control on genes was then performed. Data sets were filtered on Flags and Expression. Gene spots were removed from the analysis when:
    - The signal was not good enough. This step removed genes where expression was in the noise level i.e. there was no expression of the genes.
    - The signal was present or marginal for less than 3 microarray slides i.e values had to appear in at least half of the total number of microarray slides.
    - If necessary, data were then filtered on expression to remove unchanged genes. Upper cut off =100, lower = 20. This removes the 0-20th percentile of data.
- Statistics were then performed.
    - A paired t-test on the strain/genotype parameter was performed on the data to give a list of significant differentially expressed genes and their p-values. The null hypothesis assumed that the difference between the two observations was zero. This test was chosen as we wanted to compare differences in gene expression between a wild type (wt) strain and a deletion (ko) strain. If the ratio between the wt gene and ko gene was 1 then there was no change; $\log_2(1) = 0$.

APPENDIX IV. MICROARRAY RESULTS

Genes up-regulated upon *Rv1747* deletion (≥ 2 Fold, p-value less than 0.05)

Gene Number	Gene Name	Gene Product	Fold Up-regulated	p-value
Rv0005	*gyrB*	DNA gyrase (Subunit B) gyrB	2.0	0.001
Rv0046c	*ino1*	Myo-insositol-1-phosphate synthase ino1	2.0	3.3E-04
Rv0047c	*Rv0047c*	Conserved hypothetical protein	2.2	4.1E-05
Rv0341	*iniB*	Isoniazid inductible gene protein iniB	3.8	2.4E-04
Rv0342	*iniA*	Isoniazid inductible gene protein iniA	3.2	8.E-05
Rv0511	*hemD*	Probable uroporphyrin-III C-methyltransferase hemD	2.0	6.4E-04
Rv0676c	*mmpL5*	Probable conserved transmembrane protein mmpL5	2.0	0.009
Rv0822c	*Rv0822c*	Conserved hypothetical protein	2.0	4.0E-05
Rv0823c	*Rv0823c*	Possible transcriptional regulatory protein	2.0	1.0E-04
Rv2007c	*fdxA*	Probbale ferredoxin fdxA	2.1	0.005
Rv2396	*PE_PGRS41*	PE-PGRS family protein	2.0	7.0E-04
Rv2590	*fadD9*	Probable fatty acid coA ligase fadD9	2.0	2.0E-04
Rv3140	*fadE23*	Probable acyl-coA dehydrogenase fadE23	2.1	0.001
Rv3854c	*ethA*	Monooxygenase ethA	2.0	0.003

Genes down-regulated upon *Rv1747* deletion (≥ 2 Fold, p-value less than 0.05)

Gene Number	Gene Name	Gene Product	Fold down-regulated	p-value
CDC1551: MT0036	*CDC1551: MT0036*	Hypothetical protein	2.0	7.3E-04
CDC1551: MT2291	*CDC1551: MT2291*	Hypothetical protein	2.1	7.1E-04
Rv0091	*mtn*	Probable bifunctional mta/sah nucleosidase mtn	2.0	0.004
Rv0320	*Rv0320*	Possible conserved exported protein	2.0	0.011
Rv0620	*galK*	Probable galactokinase galK	2.0	4.1E-04
Rv1040c	*PE8*	PE family protein	2.1	0.013
Rv1380	*pyrB*	Probable aspartate carbamoyltransferase pyrB	2.2	0.001
Rv1382	*Rv1382*	Probable export or	2.2	6.2E-04

		membrane protein		
Rv1383	*carA*	Probable carbamoyl-phosphate synthase small chain carA	2.0	3.7E-05
Rv1515c	*Rv1515c*	Conserved hypothetical protein	2.0	5.1E-04
Rv1747	*Rv1747*	Probably conserved transmembrane ATP-binding protein ABC transporter	29.2	8.9E-06
Rv1756c	*Rv1756c*	Putative transposase	2.2	0.006
Rv1999c	*Rv1999c*	Probable conserved integral membrane protein	2.2	0.005
Rv2106	*Rv2106*	Probable transposase	2.1	0.016
Rv2265	*Rv2265*	Possible conserved integral membrane protein	2.1	0.006
Rv2268c	*cyp128*	Probable cytochrome P450 128 cyp128	2.1	0.005
Rv2415c	*Rv2415c*	Conserved hypothetical protein	2.3	7.9E-04
Rv2528c	*mrr*	Probable restriction system protein mrr	2.1	1.9E-04
Rv2577	*Rv2577*	Conserved hypothetical protein	2.1	0.016
Rv2812	*Rv2812*	Probable transposase	2.0	0.006
Rv2814c	*Rv2814c*	Probable transposase	2.2	0.012
Rv2874	*dipZ*	Possible integral membrane C-type cytochrome biogenesis protein dipZ	2.0	0.015
Rv3326	*Rv3326*	Probable transposase	2.0	0.016
Rv3380c	*Rv3380c*	Probable transposase	2.0	0.016
Rv3475	*Rv3475*	Probable transposase	2.1	0.014

Genes up-regulated upon *pknF* deletion (≥ 1.5 Fold, p-value less than 0.05)

Gene Number	Gene Name	Gene Product	Fold Up-regulated	p-value
CDC1551: MT3972.1	*CDC1551: MT3972.1*	Hypothetical protein	1.8	0.004
Rv0175	*Rv0175*	Probable conserved mce associated membrane protein	1.5	0.003
Rv0341	*iniB*	Isoniazid inductible gene protein iniB	1.8	0.001
Rv0831c	*Rv0831c*	Conserved hypothetical protein	1.5	0.003
Rv1738	*Rv1738*	Conserved hypothetical protein	1.6	0.005

Rv2007c	*fdxA*	Probable ferredoxin fdxA	1.8	0.023
Rv3727	*Rv3727*	Possible oxidoreductase	1.5	0.008
Rv3728	*Rv3728*	Probable conserved two-domain membrane protein	1.5	0.007
Rv3842c	*glpQ1*	Probable glycerophosphoryl diester phosphodiesterase glpQ1	1.9	0.006
Rv3850	*Rv3850*	Conserved hypothetical protein	1.7	0.002
Rv3854c	*ethA*	Monooxygenase ethA	1.7	0.006
Rv3864	*espE*	Esx-1 secretion associated protein espE	1.6	0.003

Genes down-regulated upon *pknF* deletion (≥ 2.0 Fold, p-value less than 0.05)

Gene Number	Gene Name	Gene Product	Fold Down-regulated	p-value
MbAF212297-1404	*Mb1404*	Probable conserved membrane protein	2.4	0.006
Rv0796	*Rv0796*	Putative transposase for insertion sequence element IS6110	2.0	0.036
Rv1370c	*Rv1370c*	Putative transposase for insertion sequence element IS6110	2.0	0.042
Rv1371	*Rv1371*	Probable conserved membrane protein	2.4	0.042
Rv1372	*Rv1372*	Conserved hypothetical protein	2.2	0.037
Rv2106	*Rv2106*	Probable transposase	2.0	0.029
Rv2167c	*Rv2167c*	Probable transposase	2.0	0.016
Rv2279	*Rv2279*	Probable transposase	2.0	0.030
Rv2480c	*Rv2480c*	Possible transposase for insertion sequence element IS6110	2.4	0.003
Rv2515c	*Rv2515c*	Conserved hypothetical protein	2.0	0.034
Rv2815c	*Rv2815c*	Probable transposase	2.0	0.043
Rv3640c	*Rv3640c*	Probable transposase	2.0	0.036

Genes down-regulated upon *pknF* deletion (≥ 1.5 Fold, p-value less than 0.05)

Gene Number	Gene Name	Gene Product	Fold Down-regulated	p-value
CDC1551: Mt1849.1	*CDC1551: Mt1849.1*	Hypothetical protein	1.7	0.012
CDC1551: Mt1945	*CDC1551: Mt1945*	Hypothetical protein	1.7	0.043
CDC155: Mt3269	*CDC155: Mt3269*	Hypothetical protein	1.7	0.029
CDC1551: Mt3718.1	*CDC1551: Mt3718.1*	Hypothetical protein	1.8	0.008
CDC155: Mt3767.1	*CDC155: Mt3767.1*	Hypothetical protein	1.7	0.013
Rv0071	*Rv0071*	Possible maturase	1.6	0.021
Rv0322	*udgA*	Probable udp-glucose 6-dehydrogenase udgA	1.7	0.024
Rv0520	*Rv0520*	Possible methyltransferase	1.5	0.012
Rv0584	*Rv0584*	Possible conserved exported protein	1.6	0.040
Rv0797	*Rv0797*	Putative transposase for insertion sequence element IS1547	1.6	0.047
Rv1181	*pks4*	Probable polyketide beta-ketoacyl synthase pks4	1.7	0.040
Rv1369c	*Rv1369c*	Probable transposase	1.9	0.027
Rv1552	*frdA*	Probable fumarate reductase frdA	1.6	0.019
Rv1735c	*Rv1735c*	Hypothetical membrane protein	1.7	0.023
Rv1756c	*Rv1756c*	Putative transposase	1.9	0.033
Rv1758	*cut1*	Probable cutinase cut1	1.7	0.007
Rv1762c	*Rv1762c*	Hypothetical protein	1.5	0.028
Rv1764	*Rv1764*	Putative transposase	1.9	0.034
Rv2168c	*Rv2168c*	Putative transposase for insertion sequence element IS6110	1.9	0.023
Rv2278	*Rv2278*	Putative transposase for insertion sequence element IS6110	1.9	0.038
Rv2317	*uspB*	Probable sugar-transport integral membrane protein ABC transporter uspB	1.9	0.040
Rv2319c	*Rv2319c*	Hypothetical protein	1.7	0.020
Rv2320c	*rocE*	Probable cationic amino acid transport integral membrane protein rocE	1.6	0.029
Rv2354	*Rv2354*	Probable transposase for insertion sequence element	1.6	0.049

		IS6110		
Rv2355	*Rv2355*	Probable transposase	1.9	0.027
Rv2577	*Rv2577*	Conserved hypothetical protein	1.7	0.034
Rv2648	*Rv2648*	Probable transposase for insertion sequence element IS6110	1.8	0.042
Rv2649	*Rv2649*	Probable transposase for insertion sequence element IS6110	1.9	0.031
Rv2813	*Rv2813*	Conserved hypothetical protein	1.5	0.015
Rv2814c	*Rv2814c*	Probable transposase	1.9	0.032
Rv3162c	*Rv3162c*	Possible integral membrane protein	1.5	0.005
Rv3163c	*Rv3163c*	Possible conserved secreted protein	1.5	0.045
Rv3185	*Rv3185*	Probable transposase	1.9	0.027
Rv3187	*Rv3187*	Probable transposase	1.8	0.044
Rv3189	*Rv3189*	Conserved hypothetical protein	1.7	0.030
Rv3195	*Rv3195*	Conserved hypothetical protein	1.6	0.025
Rv3324c	*moaC3*	Probable molybdenum cofactor biosynthesis protein C moaC3	1.5	0.047
Rv3325	*Rv3325*	Probable transposase for insertion sequence element IS6110	1.9	0.019
Rv3326	*Rv3326*	Probable transposase	1.9	0.028
Rv3380c	*Rv3380*	Probable transposase	1.9	0.027
Rv3473c	*bpoA*	Possible peroxidise bpoA	1.7	0.041
Rv3474	*Rv3474*	Possible transposase for insertion sequence element IS6110	1.9	0.038
Rv3475	*Rv3475*	Possible transposase for insertion sequence element IS6110	1.9	0.035
Rv3613c	*Rv3613c*	Hypothetical protein	1.8	7.0E-04
Rv3636	*Rv3636*	Possible transposase	1.6	0.027
Rv3637	*Rv3637*	Possible transposase	1.8	0.033
Rv3665c	*Rv3665c*	Porbable dipeptide-transport integral membrane protein ABC transporter dppB	1.6	0.027
Rv3752c	*Rv3752c*	Possible cytidine/deoxycytidylate deaminase	1.6	0.006